

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 18:54:25 ; Search time 48 Seconds  
(without alignments)  
3211.895 Million cell updates/sec

Title: US-09-974-973-2

Perfect score: 5865

Sequence: 1 MTAITLGLLLGIITLVST.....RVVPAATKVEGGDLIVVVS 1157

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

- 1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5865	100.0	1157	AAU98050	Corynebacterium fe
2	5779	98.5	1140	AAU98052	Corynebacterium mu
3	5764	98.3	1141	AAU01436	Pyruvate carboxyla
4	5759	98.2	1140	AAU90511	C glutamicum prote
5	5759	98.2	1140	AAU67129	Corynebacterium gl
6	5759	98.2	1140	AAU98053	Corynebacterium wi
7	5753	98.1	1140	AAU93971	C. glutamicum pyru
8	5751	98.1	1140	AAU93249	C glutamicum prote
9	5299.5	90.4	1139	AAU83180	Corynebacterium th
10	2678	45.7	532	AAU79302	Corynebacterium gl

11	2530	43.1	1148	22	AAU00511	Bacillus subtilis
12	2494	42.5	1146	23	ABU47612	Listeria monocytog
13	2489	42.4	1147	22	AAU33972	Staphylococcus aur
14	2465	42.0	1151	22	AGS2677	S. epidermidis ope
15	2465	42.0	1154	23	ABP36583	Staphylococcus epi
16	2460.5	42.0	1142	22	AAU35213	Enterococcus faeca
17	2441	41.6	1181	22	ABU58211	Drosophila melanog
18	2441	41.6	1181	22	ABU66604	Drosophila melanog
19	2441	41.6	1181	22	ABU66605	Drosophila melanog
20	2433.5	41.5	1196	22	ABU67309	Drosophila melanog
21	2412	41.1	1136	23	ABU53980	Lactococcus lactis
22	2336	39.8	461	22	AAU79303	Corynebacterium gl
23	2300	39.2	1073	22	AAU36768	Staphylococcus aur
24	1546.5	26.4	320	22	AAU79300	Corynebacterium gl
25	1546.5	26.4	320	22	AAU79301	Corynebacterium gl
26	1371	23.4	272	22	AAU79298	Corynebacterium gl
27	1371	23.4	272	22	AAU79299	Corynebacterium gl
28	1032.5	17.6	447	17	AAU05207	Anabaena biotin bi
29	1032.5	17.6	447	19	AAU70402	Anabaena biotin ca
30	1032.5	17.6	448	15	AAU51080	Biotin carboxylase
31	1005.5	17.1	453	15	AAU51083	fabG gene encoding
32	1005.5	17.1	453	17	AAU05208	Synechococcus biot
33	1005.5	17.1	453	19	AAU70403	Synechococcus biot
34	980	16.7	425	21	AAU11663	A. vitis hyperpens
35	979.5	16.7	471	22	AAU33719	Pseudomonas aerugi
36	956	16.3	449	22	AAU38292	Salmonella typhi c
37	954.5	16.3	455	23	ABU54091	Lactococcus lactis
38	950	16.2	448	22	AAU35567	Haemophilus influe
39	947.5	16.2	456	23	ABU28017	Streptococcus poly
40	940	16.0	449	16	AAU66742	Biotin-carboxylase
41	940	16.0	449	22	AAU34737	E. coli cellular p
42	937.5	16.0	454	20	AAU34779	Chlamydia pneumoni
43	936.5	16.0	455	22	AAU38024	Streptococcus pneu
44	935.5	16.0	455	22	AAU37771	Streptococcus pneu
45	933.5	15.9	455	22	AAU01054	CPE 57 protein seq

ALIGNMENTS

RESULT 1

AAU98050  
ID AAU98050 standard; Protein; 1157 AA.  
XX  
AC AAU98050;  
XX  
DT 27-AUG-2002 (first entry)  
XX  
DE Corynebacterium feedback-resistant pyruvate carboxylase enzyme.  
XX  
FE Feedback-resistant; pyruvate carboxylase; enzyme;  
KW aspartic acid feedback inhibition resistant.  
XX  
OS Corynebacterium glutamicum.  
XX  
PH Key Location/Qualifiers  
FT Region 1..18  
FT FT /note= "Specifically claimed in claim 18"  
FT FT 164..176  
FT FT /note= "Specifically claimed in claim 18"  
FT FT 193..205  
FT FT /note= "Specifically claimed in claim 18"  
FT FT 217..229  
FT FT /note= "Specifically claimed in claim 18"  
FT FT 238..250  
FT FT /note= "Specifically claimed in claim 18"  
FT FT 466..478  
FT FT /note= "Specifically claimed in claim 18";  
XX WO200231158-A2.  
XX 18-APR-2002.  
XX

PF 12-OCT-2001; 2001WO-US31893.  
 XX 13-OCT-2000; 2000US-239913P.  
 XX (ARCH ) ARCHER-DANIELS MIDLAND CO.  
 PA Hanke PD;  
 XX WPI: 2002-463267/49.  
 XX N-PSDB; ABK52832.  
 DR Novel mutated, feedback resistant pyruvate carboxylase enzyme  
 PT polypeptide, useful for producing amino acids e.g. L-lysine,  
 PT L-threonine, L-glycine, L-glutamic acid, L-proline and L-methionine and  
 PT L-isoleucine  
 XX Claim 17; Fig 1; 42pp; English.  
 XX The present invention relates to a new mutated, feedback-resistant  
 CC pyruvate carboxylase enzyme. The invention is useful for producing an  
 CC amino acid (e.g. L-lys, L-thr, L-Met, L-Ile, L-Glu, L-Arg and L-Pro),  
 CC by culturing a host cell in a suitable media and separating the amino  
 CC acid from the medium. The vector of the invention is useful for  
 CC replacement of a wild-type pyruvate carboxylase gene, with a feedback  
 CC resistant pyruvate carboxylase gene, in a Corynebacterium glutamicum, by  
 CC replacing a genomic copy of the wild-type pyruvate carboxylase gene, with  
 CC a selectable marker gene through homologous recombination to form a first  
 CC recombinant strain, and replacing the selectable marker gene in the  
 CC first recombinant strain, with feedback resistant pyruvate carboxylase  
 CC gene through homologous recombination to form a second recombinant  
 CC strain, where the homologous recombination in the above steps, occurs  
 CC between the host cell and the vector. The feedback-resistant pyruvate  
 CC carboxylase enzyme is resistant to feedback inhibition from aspartic  
 CC acid. The present amino acid sequence represents the feedback-resistant  
 CC pyruvate carboxylase enzyme of the invention.  
 XX  
 SQ Sequence 1157 AA;  
 Query Match 100.0%; Score 5865; DB 23; Length 1157;  
 Best Local Similarity 100.0%; Pred. NO. 0;  
 Matches 1157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MTATLGLLLKGIITLVSTHSTSLPAFKKILVANRGEIAVRAALETGAATVAIYP 60  
 DB 1 MTATLGLLLKGIITLVSTHSTSLPAFKKILVANRGEIAVRAALETGAATVAIYP 60  
 QY 61 REDGSHFRSFASAVRIGTSGSPVKAYLDIDEIIGAACKVKADAIYPGYGFLSENAOLA 120  
 DB 61 REDGSHFRSFASAVRIGTSGSPVKAYLDIDEIIGAACKVKADAIYPGYGFLSENAOLA 120  
 QY 121 RECAENGITFTGPTPEVLDLTDGDKSRATVAAKAGLPVLAESTPSKNIDDTIVKSAEGQTY 180  
 DB 121 RECAENGITFTGPTPEVLDLTDGDKSRATVAAKAGLPVLAESTPSKNIDDTIVKSAEGQTY 180  
 QY 181 PIFVKAVAGGGGMRFRVSSPDEURKLATEASREAEAFGDSGYVERAVINPOHIEVQI 240  
 DB 181 PIFVKAVAGGGGMRFRVSSPDEURKLATEASREAEAFGDSGYVERAVINPOHIEVQI 240  
 QY 241 LGDRTGEVHHLYERDCSLQRRHQKVETAPAQHLDPELRDRIKADAVKFCISGYQAGT 300  
 DB 241 LGDRTGEVHHLYERDCSLQRRHQKVETAPAQHLDPELRDRIKADAVKFCISGYQAGT 300  
 QY 301 VEFLVDEKGNHVFIEIMNPRIQVEHTVTEVTEVDLVKAQMRLAAGATLKEGLTQDKIKT 360  
 DB 301 VEFLVDEKGNHVFIEIMNPRIQVEHTVTEVTEVDLVKAQMRLAAGATLKEGLTQDKIKT 360  
 QY 361 HCAALQCRITTTDPNNGRPDGTITAYRSPGGAGVRDGAAGLGGETTAHFDSMLVKMT 420  
 DB 361 HCAALQCRITTTDPNNGRPDGTITAYRSPGGAGVRDGAAGLGGETTAHFDSMLVKMT 420  
 QY 421 CRGSDFEFAVARAORALAEFTVSGVATNIGFLRALLREEDFTSKRIATGFGDHPHLLQA 480  
 DB 421 CRGSDFEFAVARAORALAEFTVSGVATNIGFLRALLREEDFTSKRIATGFGDHPHLLQA 480

QY 481 PPADDEQGRILDYLDADVTNPKPHGVPRKDVAAPIDKLPNIKDLPLPRGSRDLKOLGFAA 540  
 DB 481 PPADDEQGRILDYLDADVTNPKPHGVPRKDVAAPIDKLPNIKDLPLPRGSRDLKOLGFAA 540  
 QY 541 FARDLREODALAVDTTFRDAHOSLLATRVRSFALKPAAEAVAKLTPELLSVEAMGGATY 600  
 DB 541 FARDLREODALAVDTTFRDAHOSLLATRVRSFALKPAAEAVAKLTPELLSVEAMGGATY 600  
 QY 601 DYAMRFLFEDPWDRLEDELRAMPNNVNIOMLLRGRNTVGYTYPDSVCRAFVKEAASSVD 660  
 DB 601 DYAMRFLFEDPWDRLEDELRAMPNNVNIOMLLRGRNTVGYTYPDSVCRAFVKEAASSVD 660  
 QY 661 IFRIEDALNDVSQMRPAIDAVLETNTAVAEVAMAYSGDLSDPNEKLYTLDYLLKMAEIV 720  
 DB 661 IFRIEDALNDVSQMRPAIDAVLETNTAVAEVAMAYSGDLSDPNEKLYTLDYLLKMAEIV 720  
 QY 721 KSGAHILAIKDMAGLLRPAAVTKLTALRRREFDLPVHVTHTDTAGGOLATYFAAAQAGAD 780  
 DB 721 KSGAHILAIKDMAGLLRPAAVTKLTALRRREFDLPVHVTHTDTAGGOLATYFAAAQAGAD 780  
 QY 781 AVDGASAPLSGTTTSSPSLSAISVAFAHTRRDTGLSLEAVSDELPYWEAVRGLYLPFESGT 840  
 DB 781 AVDGASAPLSGTTTSSPSLSAISVAFAHTRRDTGLSLEAVSDELPYWEAVRGLYLPFESGT 840  
 QY 841 PGPTGRVYRHEIPGQSLNLRQAOTALGLADRFELIEDNYAAVNEMLRGPKTKVTPSSKVV 900  
 DB 841 PGPTGRVYRHEIPGQSLNLRQAOTALGLADRFELIEDNYAAVNEMLRGPKTKVTPSSKVV 900  
 QY 901 GDIALHLVAGVDPADRADPOKYDIPDSVIAFLRGLGNPGGWPPEPLRTWALEGRSEG 960  
 DB 901 GDIALHLVAGVDPADRADPOKYDIPDSVIAFLRGLGNPGGWPPEPLRTWALEGRSEG 960  
 QY 961 KAPLTVPEEQAHLDDADDKSKERNSLNRLFPKTEFELEHRRFRFGNTSALDDREFFYG 1020  
 DB 961 KAPLTVPEEQAHLDDADDKSKERNSLNRLFPKTEFELEHRRFRFGNTSALDDREFFYG 1020  
 QY 1021 LVEGRRTLRLPDVRTPLLRDLDAISEPDDKCMRNVNANVNGOIRPMVRDRSVESVTAT 1080  
 DB 1021 LVEGRRTLRLPDVRTPLLRDLDAISEPDDKCMRNVNANVNGOIRPMVRDRSVESVTAT 1080  
 QY 1081 AEKADSSNKGHVAAPFAGVYVTVFAGDEBKAGDAVAIIIEAMKMEATITASVDGKIERYV 1140  
 DB 1081 AEKADSSNKGHVAAPFAGVYVTVFAGDEBKAGDAVAIIIEAMKMEATITASVDGKIERYV 1140  
 QY 1141 VPAATKVEGGDLIVVVS 1157  
 DB 1141 VPAATKVEGGDLIVVVS 1157  
 RESULT 2  
 AAU98052  
 ID AAU98052 standard; Protein; 1140 AA.  
 XX  
 AC AAU98052;  
 XX  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE Corynebacterium mutant feedback-resistant pyruvate carboxylase enzyme.  
 XX  
 KW Feedback-resistant; pyruvate carboxylase; enzyme;  
 XX aspartic acid feedback inhibition resistant; mutant; muten.  
 KW  
 XX Corynebacterium glutamicum.  
 OS  
 XX Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FT Misc-difference 1 /note= "Wild-type Met substituted by Val"  
 FT Misc-difference 153 /note= "Wild-type Glu substituted by Asp"  
 FT Misc-difference 182 /note= "Wild-type Ala substituted by Ser"  
 FT

FT Misc-difference 206 /note= "Wild-type Ala substituted by Ser"  
 FT Misc-difference 227 /note= "Wild-type His substituted by Arg"  
 FT Misc-difference 455 /note= "Wild-type Ala substituted by Gly"  
 FT Region 1110..1122 /note= "Specifically claimed in claim 18"  
 FT Misc-difference 1116 /note= "Wild-type Asp substituted by Glu"  
 XX  
 PN WO200231158-A2.  
 XX 18-APR-2002.  
 XX  
 XX 12-OCT-2001; 2001WO-US31893.  
 XX  
 XX 13-OCT-2000; 2000US-239913P.  
 PR (ARCH ) ARCHER-DANIELS MIDLAND CO.  
 XX  
 XX Hanke PD;  
 XX  
 DR WPI; 2002-463267/49.  
 XX  
 XX Novel mutated, feedback resistant pyruvate carboxylase enzyme  
 PT polypeptide, useful for producing amino acids e.g. L-lysine  
 PT L-threonine, L-glycine, L-glutamic acid, L-proline and L-methionine and  
 PT L-isoleucine  
 XX  
 PS Claim 1; Page -: 42pp; English.  
 XX  
 CC The present invention relates to a new mutated, feedback-resistant  
 CC pyruvate carboxylase enzyme. The invention is useful for producing an  
 CC amino acid (e.g. L-lys, L-thr, L-Met, L-Ile, L-Glu, L-Arg and L-Pro),  
 CC by culturing a host cell in a suitable media and separating the amino  
 CC acid from the medium. The vector of the invention is useful for  
 CC replacement of a wild-type pyruvate carboxylase gene, with a feedback  
 CC resistant pyruvate carboxylase gene, in a Corynebacterium glutamicum, by  
 CC replacing a genomic copy of the wild-type pyruvate carboxylase gene with  
 CC a selectable marker gene through homologous recombination to form a first  
 CC recombination strain, and replacing the selectable marker gene in the  
 CC first recombinant strain, with feedback resistant pyruvate carboxylase  
 CC gene through homologous recombination to form a second recombinant  
 CC strain, where the homologous recombination in the above steps, occurs  
 CC between the host cell and the vector. The feedback-resistant pyruvate  
 CC carboxylase enzyme is resistant to feedback inhibition from aspartic  
 CC acid. The present amino acid sequence represents the mutant  
 CC feedback-resistant pyruvate carboxylase enzyme of the invention.  
 CC Note: The present sequence is not shown in the specification but is  
 CC derived from the wild-type feedback-resistant pyruvate carboxylase  
 CC enzyme (AAU98053) given in figure 2 of the specification.  
 XX  
 SQ Sequence 1140 AA;  
 Query Match 98.5%; Score 5779; DB 23; Length 1140;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1138; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 QY 18 VSTHTSSTLPAPKILVANRGEIAVRAALETGAATVAIYPREDRGSFHRSEAVR 77  
 Db 1 VSTHTSSTLPAPKILVANRGEIAVRAALETGAATVAIYPREDRGSFHRSEAVR 60  
 QY 78 ICTEGSPVKAYIDIDEIIGAANKVRADAIYPGYFLSENAQLARECAENGITFIPTPEV 137  
 Db 61 ICTEGSPVKAYIDIDEIIGAANKVRADAIYPGYFLSENAQLARECAENGITFIPTPEV 120  
 QY 138 LDLTGDKSRVTAANKAGLPVLAESTPSKNIDDIVKSAEGQYPIPVKAVAGGGGRGMRF 197  
 Db 121 LDLTGDKSRVTAANKAGLPVLAESTPSKNIDDIVKSAEGQYPIPVKAVAGGGGRGMRF 180  
 QY 198 VSSPDELKRLATEASREAEAFGDGSSVYVERAVINPQHIEVQILGDRTEGEVHLXERDCS 257  
 Db 198 VSSPDELKRLATEASREAEAFGDGSSVYVERAVINPQHIEVQILGDRTEGEVHLXERDCS 240

Db 181 VSSPDELKRLATEASREAEAFGDGSSVYVERAVINPQHIEVQILGDRTEGEVHLXERDCS 240  
 QY 258 LORRHOKVVEIAPAOHLDPDLRICADAVKFCRSIGYOGAGTVEFLVDEKGNHVEIWN 317  
 Db 241 LORRHOKVVEIAPAOHLDPDLRICADAVKFCRSIGYOGAGTVEFLVDEKGNHVEIWN 300  
 QY 318 PRIQVEHTVTEEVVDLVKAQMLAAGATLKEGLTQDKIKTHGAALOCRTTDPNNG 377  
 Db 301 PRIQVEHTVTEEVVDLVKAQMLAAGATLKEGLTQDKIKTHGAALOCRTTDPNNG 360  
 QY 378 FRPDGTGITAYRSPGGAGVRLDGAALGGEITAHFDSMLVKMTCRGSDETAVARAQRAL 437  
 Db 361 FRPDGTGITAYRSPGGAGVRLDGAALGGEITAHFDSMLVKMTCRGSDETAVARAQRAL 420  
 QY 438 AEFTVSGVATNIGFLRALLREEDFTSKRIATGFIGDHPHLLQAPPADDDQGRITLDYADV 497  
 Db 421 AEFTVSGVATNIGFLRALLREEDFTSKRIATGFIGDHPHLLQAPPADDDQGRITLDYADV 480  
 QY 498 TVNKPCHGVPRKDVAAPIDKLPNIKDLPLPRGSRDLKQLGPAFAFARDLREQDALAVTDIT 557  
 Db 481 TVNKPCHGVPRKDVAAPIDKLPNIKDLPLPRGSRDLKQLGPAFAFARDLREQDALAVTDIT 540  
 QY 558 FRDAHQSLLATRVRSFALKPAEAAVAKLTPELLSVEAWGATYDVAMRFLFEDPMDRLDE 617  
 Db 541 FRDAHQSLLATRVRSFALKPAEAAVAKLTPELLSVEAWGATYDVAMRFLFEDPMDRLDE 600  
 QY 618 LREAMPNVNIOMLLRCRNTVGTPYDPSVCRAPVKEAASGVDFIRIFDALNDVSOMRPA 677  
 Db 601 LREAMPNVNIOMLLRCRNTVGTPYDPSVCRAPVKEAASGVDFIRIFDALNDVSOMRPA 660  
 QY 678 IDAVLETNTAVAEVAMAYSGDLPNEKLYTLDDYILKMAEIEVKSGAHLIAIKDMAGLLR 737  
 Db 661 IDAVLETNTAVAEVAMAYSGDLPNEKLYTLDDYILKMAEIEVKSGAHLIAIKDMAGLLR 720  
 QY 738 PAAVTKLVTLARREEDLPVHVHTHDTAGQLATYFAAAQAGADAVDGAAPLSGTTSPQS 797  
 Db 721 PAAVTKLVTLARREEDLPVHVHTHDTAGQLATYFAAAQAGADAVDGAAPLSGTTSPQS 780  
 QY 798 LSAIAVAAFAHTRDTGLSLSEAVSDLEPYWEAVRGLVLPFESGTPGPTGRVYRHEIPGGOL 857  
 Db 781 LSAIAVAAFAHTRDTGLSLSEAVSDLEPYWEAVRGLVLPFESGTPGPTGRVYRHEIPGGOL 840  
 QY 858 SNLRAQATGALGLADREFELIEDNYAAVNEMLGRPTKVTSPSSKVVGDIALHLVGAGVDPADF 917  
 Db 841 SNLRAQATGALGLADREFELIEDNYAAVNEMLGRPTKVTSPSSKVVGDIALHLVGAGVDPADF 900  
 QY 918 AADPKYIDIPDSVIAFLRGLGNPPGWPPEPLTRALEGRSEKAPLTVPEEEQAHLDA 977  
 Db 901 AADPKYIDIPDSVIAFLRGLGNPPGWPPEPLTRALEGRSEKAPLTVPEEEQAHLDA 960  
 QY 978 DSKERNLSNRLLPKPTTEEFLEHRRRFGNTSALDDREFFYGLVGEVRETILRLPDVPTP 1037  
 Db 961 DSKERNLSNRLLPKPTTEEFLEHRRRFGNTSALDDREFFYGLVGEVRETILRLPDVPTP 1020  
 QY 1038 LLVRLDAISEPDDKGMNRNVANVNGQIRPMRVDRSVEVSTATAEKADSSNKGHVAAPFA 1097  
 Db 1021 LLVRLDAISEPDDKGMNRNVANVNGQIRPMRVDRSVEVSTATAEKADSSNKGHVAAPFA 1080  
 QY 1098 GVVTVTVAEGDEVKAGDAVAIIEAMKMEATITASVDGCKTERVVVPAATKVEGDLIVVVS 1157  
 Db 1081 GVVTVTVAEGDEVKAGDAVAIIEAMKMEATITASVDGCKTERVVVPAATKVEGDLIVVVS 1140  
 RESULT 3  
 ID AAB01436  
 XX AAB01436 standard; Protein; 1141 AA.  
 AC AAB01436;  
 DT 20-OCT-2000 (first entry)  
 DE Pyruvate carboxylase of C. glutamicum.  
 XX











Db 121 LDLTGDKSRVATAAKKAGLPVLAESTPSKNIIDEIVKSAEGQTYPIFVKAVAGGGGRMRF 180  
 Qy 198 VSSPDELRLKRLATEASREAEAFDGSVYVERAVINPOHIEVQILGDRGTGEVVLHYERDCS 257  
 Db 181 VASPDRLKRLATEASREAEAFDGSVYVERAVINPOHIEVQILGDRGTGEVVLHYERDCS 240  
 Qy 258 LORRHQKVETAPQAHLDPQLDRICADAVKFCRSIGYQAGTVEFLVDEKGNHVEIEN 317  
 Db 241 LORRHQKVETAPQAHLDPQLDRICADAVKFCRSIGYQAGTVEFLVDEKGNHVEIEN 300  
 Qy 318 PRIOVEHTVTEVTEVDLVKRAOMRLAAGATLKGELGTQDKIKTHGAALQCRITTEPNNG 377  
 Db 301 PRIOVEHTVTEVTEVDLVKRAOMRLAAGATLKGELGTQDKIKTHGAALQCRITTEPNNG 360  
 Qy 378 PRDTGITAYRSPGGAGVRLDGAQJGGEITAHFDSMLVKMTCRGSDFETAVARAQAL 437  
 Db 361 PRDTGITAYRSPGGAGVRLDGAQJGGEITAHFDSMLVKMTCRGSDFETAVARAQAL 420  
 Qy 438 AEFTVSGVATNIGFLRALLREEDFTSKRIATGFTIGDRPHLLQAPPADDEQGRILDYLDV 497  
 Db 421 AEFTVSGVATNIGFLRALLREEDFTSKRIATGFTIADPHLLQAPPADDEQGRILDYLDV 480  
 Qy 498 TVNKPBGVVRPKDVAAPIDKLPNFKDLPLPRGSRDLKQLGPAFAFARDLREQDALAVTDTT 557  
 Db 481 TVNKPBGVVRPKDVAAPIDKLPNFKDLPLPRGSRDLKQLGPAFAFARDLREQDALAVTDTT 540  
 Qy 558 FRDAHQSLLATRVRSFALKPAEAEVAKLTPELLSVEAWGATYDVAMRFLFEDPWRDLDE 617  
 Db 541 FRDAHQSLLATRVRSFALKPAEAEVAKLTPELLSVEAWGATYDVAMRFLFEDPWRDLDE 600  
 Qy 618 LREAMPNVNIOMLRGNTVGYTPYPSVCRFAFVKEAASGVDFIRFDALNDVDSOMRPA 677  
 Db 601 LREAMPNVNIOMLRGNTVGYTPYPSVCRFAFVKEAASGVDFIRFDALNDVDSOMRPA 660  
 Qy 678 IDAVLENTVAEYAMAYSGDLSPNEKLYLDYLYLKMAEIVKSGAHTLAKDMAGLLR 737  
 Db 661 IDAVLENTVAEYAMAYSGDLSPNEKLYLDYLYLKMAEIVKSGAHTLAKDMAGLLR 720  
 Qy 738 PAAVTKLVTLRRFEDLPVHVHHTDAGQOLATYFAAQAQADAVDGSAPLSTGTSQPS 797  
 Db 721 PAAVTKLVTLRRFEDLPVHVHHTDAGQOLATYFAAQAQADAVDGSAPLSTGTSQPS 780  
 Qy 798 LSAIVAAFAHTRRTDGLSLAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGGOL 857  
 Db 781 LSAIVAAFAHTRRTDGLSLAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGGOL 840  
 Qy 858 SNLRAQATAGLADRFELIEDNYAAVNEMLGRPTKVTTPSSKVYVGDALHLVAGVDPADF 917  
 Db 841 SNLRAQATAGLADRFELIEDNYAAVNEMLGRPTKVTTPSSKVYVGDALHLVAGVDPADF 900  
 Qy 918 AADPKYDIPDSVIAFLRGLNPPGWPPEPLRTRALEGRSECKAPLTPVEPEEQAHLLA 977  
 Db 901 AADPKYDIPDSVIAFLRGLNPPGWPPEPLRTRALEGRSECKAPLTPVEPEEQAHLLA 960  
 Qy 978 DSKERNLSNRLLPKPTBEFLHRRFRFGNTSALDDREFFYGLVGRFETLRLPDVTRP 1037  
 Db 961 DSKERNLSNRLLPKPTBEFLHRRFRFGNTSALDDREFFYGLVGRFETLRLPDVTRP 1020  
 Qy 1038 LVLRLDAISEPDDKGMNVANVNGQIRPMVRDRSVESVTATAEKADSNKGHVAAPPA 1097  
 Db 1021 LVLRLDAISEPDDKGMNVANVNGQIRPMVRDRSVESVTATAEKADSNKGHVAAPPA 1080  
 Qy 1098 GVVTVTVAEGDEVKAGDAVALIEMKMEATITASVDGKIIEVVVPAATKVEGGDLIVVVS 1157  
 Db 1081 GVVTVTVAEGDEVKAGDAVALIEMKMEATITASVDGKIIEVVVPAATKVEGGDLIVVVS 1140

RESULT 8  
 ID AAG93249 standard; Protein: 1140 AA.  
 XX  
 AC AAG93249;  
 XX

DT 26-SEP-2001 (first entry)  
 XX C glutamicum protein fragment mutant P458S.  
 DE  
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis; mutant; muten.  
 XX Corynebacterium glutamicum.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 458 /note= "wild-type Pro substituted by Ser"  
 FT  
 XX  
 PN EPI108790-A2.  
 XX  
 XX 20-JUN-2001.  
 PD  
 XX 18-DEC-2000; 2000EP-0127688.  
 PF  
 XX 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA  
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX  
 XX WPI; 2001-376931/40.  
 DR  
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 XX  
 PS Claim 43: Page -: 246pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a mutant protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 SQ Sequence 1140 AA;  
 Query Match 98.1%; Score 5751; DB 22; Length 1140;  
 Best Local Similarity 99.3%; Pred No. 0;  
 Matches 1132; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
 Qy 18 VSTHTSSILPAEKKILVANRGEIARAFRAALTAETGAATVAIYIPREDRGSFHRSEAVR 77  
 Db 1 MSHHTSSILPAEKKILVANRGEIARAFRAALTAETGAATVAIYIPREDRGSFHRSEAVR 60  
 Qy 78 IGTGSPVKAYLIDDEIIGAARKKADAIYPGYGLSNAQLARECAENGITFTIGTPEV 137  
 Db 61 IGTGSPVKAYLIDDEIIGAARKKADAIYPGYGLSNAQLARECAENGITFTIGTPEV 120  
 Qy 138 LDLTGDKSRVATAAKKAGLPVLAESTPSKNIIDDITVKSABGQTYPIFVKAVAGGGGRMRF 197  
 Db 121 LDLTGDKSRVATAAKKAGLPVLAESTPSKNIIDEIVKSAEGQTYPIFVKAVAGGGGRMRF 180  
 Qy 198 VSSPDELRLKRLATEASREAEAFDGSVYVERAVINPOHIEVQILGDRGTGEVVLHYERDCS 257  
 Db 181 VASPDRLKRLATEASREAEAFDGSVYVERAVINPOHIEVQILGDRGTGEVVLHYERDCS 240





Db 1 VTAITLGGILLKGIITLVSTHTSSTLPAPFKKILVANRGEIAVRAFRALLETGAATVAIYP 60  
QY 61 REDRGSFHRSEAFSEAVRIGTEGSPVKAYLDIDELIICAARKVKADAIYPGYFLSENAQLA 120  
Db 61 REDRGSFHRSEAFSEAVRIGTEGSPVKAYLDIDELIICAARKVKADAIYPGYFLSENAQLA 120  
QY 121 RECAENGITFIGTPEVLDLTGDKSRVTAAKKAGLPVLAESTPSSKNIDDIIVKSAEGQTY 180  
Db 121 RECAENGITFIGTPEVLDLTGDKSRVTAAKKAGLPVLAESTPSSKNIDDIIVKSAEGQTY 180  
QY 181 PIFVKAVAGGGGGRMFVSPDELRLKLAETASREAEAAFGDGSVYVERAVINPOHTEVOI 240  
Db 181 PIFVKAVAGGGGGRMFVSPDELRLKLAETASREAEAAFGDGSVYVERAVINPOHTEVOI 240  
QY 241 LGDRTGEVHLVERDCSLORRHQKVVEIAPAOHLDPDELDRICADAVKFCRSIGYGAGT 300  
Db 241 LGDHTGEVHLVERDCSLORRHQKVVEIAPAOHLDPDELDRICADAVKFCRSIGYGAGT 300  
QY 301 VFELVDEKGNHVFIEIENPRIQVHEVTVEEVDLVKAQMRLAAGATLKELGLTQDKIKT 360  
Db 301 VFELVDEKGNHVFIEIENPRIQVHEVTVEEVDLVKAQMRLAAGATLKELGLTQDKIKT 360  
QY 361 HGAALQCRITTEDPNNNGFRPDGTITAYRSPGGAGVRLDGAALGGEITAHFDSMLVKMT 420  
Db 361 HGAALQCRITTEDPNNNGFRPDGTITAYRSPGGAGVRLDGAALGGEITAHFDSMLVKMT 420  
QY 421 CRGSDPETAARAQALAEFTVSGVATNIGFLRALLREEDFTSKRIATGFIGDHPHLLQA 480  
Db 421 CRGSDPETAARAQALAEFTVSGVATNIGFLRALLREEDFTSKRIATGFIGDHPHLLQA 480  
QY 481 PPADDEQGRILDYLDADVTYNKPHGVPRKDVAAPIDKLPNIKDLPLPRGSRDR 532  
Db 481 PPADDEQGRILDYLDADVTYNKPHGVPRKDVAAPIDKLPNIKDLPLPRGSRDR 532  
RESULT 11  
AAU00511  
ID AAU00511 standard; Protein; 1148 AA.  
AC AAU00511;  
XX  
DT 07-SEP-2001 (first entry)  
XX  
DE Bacillus subtilis pyruvate carboxylase enzyme A.  
KW Pyruvate carboxylase A; pyCA: oxaloacetate; Escherichia;  
KW L-amino acid production; fermentation.  
XX  
OS Bacillus subtilis strain 168.  
XX  
PN EP1092776-A1.  
XX  
PD 18-APR-2001.  
XX  
PF 05-OCT-2000; 2000EP-0121763.  
XX  
PR 14-OCT-1999; 99RU-0121636.  
XX  
PA (AJIN ) AJINOMOTO CO INC.  
XX  
PI Gussyatiner MM, Kozlov YI, Ptitsyn LR, Altman IB, Voroshilova EB;  
PI Iomantas YAV, Yampolskaya TA;  
XX  
DR WPI: 2001-309819/33.  
DR N-PSDB; AAS01509.  
XX  
PT New bacterium from the genus Escherichia containing a gene encoding for  
PT pyruvate carboxylase useful for producing higher concentrations of  
XX L-amino acids  
PS Claim 4; Page 17-21; 28pp; English.  
XX  
CC The present sequence represents Bacillus subtilis pyruvate

CC carboxylase enzyme A. The pyCA gene encodes for pyruvate carboxylase  
CC which catalyses the carboxylation of pyruvate to form oxaloacetate.  
CC Transformation of the Bacillus subtilis pyCA gene into a bacterium  
CC belonging to the genus Escherichia results in the bacterium showing  
CC L-amino acid productivity. The invention provides a novel method for  
CC producing an L-amino acid by fermentation. The method involves  
CC cultivating the bacterium in a medium and producing and accumulating  
CC the L-amino acid in the medium. The new bacterium harbouring the gene  
CC coding for an enzyme having pyruvate carboxylase activity is useful  
CC for producing higher concentrations of L-amino acids in vitro than  
CC prior art.  
XX  
SQ Sequence 1148 AA;  
Query Match 43.1%; Score 2530; DB 22; Length 1148;  
Best Local Similarity 46.0%; Pred. No. 2.2e-179;  
Matches 533; Conservative 188; Mismatches 390; Indels 48; Gaps 13;  
QY 28 AFKKILVANRGEIAVRAFRALLETGAATVAIYPREDRGSFHRSEAFSEAVRIGTEGSPVKA 87  
Db 5 SIQKVLVANRGEIAVIRFRACTELNIRTVAVYSKEDSGSYHRYKADAEVLVGSKKPIDA 64  
QY 88 YLDIDELIICAARKVKADAIYPGYFLSENAQLARECAENGITFIGTPEVLDLTGDKSRA 147  
Db 65 YLDIEGIIIDIAKRNKVDALHPGYGFLSENIHFARCEEGEIVFIGPKSEHLDMFGDKVKA 124  
QY 148 VTAARKAGLPVLAEST-PSKNIDDIIVKSAEGQTYPIFVKAVAGGGGGRMFVSSPDELK 206  
Db 125 RQAEKAGIPVPGSDGPAETLEAVEQFGAANGYPIIIKASLGGGGGGRMIVRSESEVKE 184  
QY 207 LATEAREAEAAFGDGSVYVERAVINPOHIEVQILGDRTEGVVHLVERDCSLQRRHQKV 266  
Db 185 AYERAKSEAKAAGFNDDEVYVEKLIENPKHIEVQVIGDKQGNVHLFERDCSVQRRHQVI 244  
QY 267 ETAPAOHLDPDELDRICADAVKFCRSIGYGAGTVEFLVDEKGNHVFIEIENPRIQVHEV 326  
Db 245 EVAPSVLSPELRDQICEAAALAKNVNINAGTVLEFLV-ANNEFYFIEVNPVQVHEITI 303  
QY 327 TEEVTEVDLVKAQMRLAAGATL-KELGLTQDK-IKTHGAALQCRITTEDPNNNGFRPDGT 383  
Db 304 TEMITGVDIVQIILVAQGHLSHKVNIPKQNDIETIGYAIQSRVTTEDPQNDPMDPTG 363  
QY 384 TITAYRSPGGAGVRLD-GAALGGEITAHFDSMLVKMTCRGSDPETAARAQALAEFTV 442  
Db 364 KIMAYRSGGGFVRLDTGNSFGAVITPYDLSLVKLTWALTFEQAAKMYRNLOEFRI 423  
QY 443 SGVATNIGFLRALLREEDFTSKRIATGFIGDHPHLLQAPADDEQGRILDYLDADVTYN-- 500  
Db 424 RGIKTNIPELVNAKHEKELTQGYDTSFIDTTPELNFPKQKDRGTMKLTITGNVTNGF 483  
QY 501 -----KPHGVPRKDVAAPIDKLPNIKDLPLPRGSRDRILKOLGPAFAAFARDLREQDALAV 553  
Db 484 PGIGKKEKPAFDKPLGVKVDVDDQP-----ARGTKQILDEKGAEGLANVKEQKSVLL 536  
QY 554 TDTTFRDAHQSLLATRVRSFALKPAEAVAKLTPELLSVSEAMGGATYDVAARFLFEDPDND 613  
Db 537 TDTTFRDAHQSLLATRIKSHDLKKTANPTAALWPELFSMEWMGGATFDVAYRFLKEDPNK 596  
QY 614 RLDELREAMPNVNIQMLLRGNTVGYTPYDSCVAFVKEAAASSGVVDIFRFDALNDVSO 673  
Db 597 RLEDLRKEVPNTLQFQMLRRSSNAVGYTNPVNIKEFVKQSAQSGIDVFIRFDSLNNWKG 656  
QY 674 MRPAIDAVLETNTVAEAVAMAYSGDLSDPNEKLYTLDDYLLKMAEEIVKSGAHTLAKDMA 733  
Db 657 MTLAIDAVRDTG-KVAEAAICTYTGDIIDKRNTRYDLYATYSMAKELEAGAHILGLKDMA 715  
QY 734 GLLRPAAVTKLVTALRREFDLPVHVHTHTAGGQATYFAAAQAGADADVASAPLSTGT 793  
Db 716 GLLKQAAVELVSALKETIDIPVHLHTDTSNGITMYAKAVEAGVDIIDVAVSSMAGLT 775  
QY 794 SQPSLSAIVAAFAHTRRDTGLSLEAVSDLEPYWEAVRGLYLPFESGTPGTGRVYRHEIP 853  
Db 776 SQPSASGEYHAMEGNDRRRPMNVQGVELLQYVESVRYKYYSEFESGMSKSPHTEIYEHMP 835









Db 705 PHILAIDMAGLLKPKAAAYELIGELKAADLPILHLDHTSGNLLTYKQADAGVDIID 764  
 Qy 784 GASAPLSGTSQSLSAIVAAFAHTRDRTGLSEAVSDLEPYWEAVRGLYLPFESGTPG 843  
 Db 765 TAVASMSGLTSQPSANSLYXALNGFPFHLRTDIEGMSUSHYVSTVYTSFESDIKSP 824  
 Qy 844 TGRVYRHEIFGGQSLNRAQATGALGLADREFELIEDNYAAVNEMLGRTKVTTPSSKVVGD 903  
 Db 825 NTELYQHEMFGGQYSLNSQAKSLGLGERDEVDKMYRRVNFGLGDIVKVTTPSSKVVGD 884  
 Qy 904 ALHLVGVGVPADPAADPQRYDIPDSVIAFLRGLGNPFGWPEPLRTRALEGRSEKAP 963  
 Db 885 ALYVQNDLDEQSVITDGYKLFPPESVSPFKGEIGOPVNGFNKDLQAVILKQOE----A 940  
 Qy 964 LTVPEEQAHLDADDSKE-----RRNSLRLLFPKPTPEFLHRRRFGNTSAL 1012  
 Db 941 LTRAPGEYLPVDFEKVRELLEEEQGPVTEQDIISVLYPKVYEQYIQTRNOYGNLSLL 1000  
 Qy 1013 DDREFFGLVGEVRETLRLPDVTRPLILVRLDAISEDDKMRNVVANVQTRPMVRDR 1072  
 Db 1001 DTPFTFFGMRNGETVEIEI-DKGRLLIRLDETISEPDENGNTIYYAMNGOARITIKDE 1059  
 Qy 1073 SVESVTATAEKADSSNGHVAAPFAGVVT-VTVAGDEGVKAGDAVAIIEMAKMEATITAS 1131  
 Db 1060 NVHTNANVKPKAKSNPSHIGAQMPGSVTEVKVSGVETKANOPLLITEAMKMETTIOAP 1119  
 Qy 1132 VDGKIERVVVPAATKVEGGDLIVV 1156  
 Db 1120 FDGVIKQVTVNGDGIATGDLIIIEI 1144  
 RESULT 14  
 AAG82677  
 ID AAG82677 standard; Protein; 1151 AA.  
 XX AC AAG82677;  
 XX DT 03-SEP-2001 (first entry)  
 XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:2448.  
 XX KW Staphylococcus epidermidis S81 strain; infection; diagnosis;  
 XX KW vaccination; endocarditis.  
 XX OS Staphylococcus epidermidis.  
 XX PN W0200134809-A2.  
 XX PD 17-MAY-2001.  
 XX PF 09-NOV-2000; 2000WO-US30782.  
 XX PR 09-NOV-1999; 99US-0164258.  
 XX PA (GLAX ) GLAXO GROUP LTD.  
 XX PI KImmerly WJ;  
 XX PI WPI: 2001-316495/33.  
 XX DR N-PSDB; AAH53527.  
 XX PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,  
 XX PT useful for vaccinating against infections, e.g. endocarditis -  
 XX PS Claim 18; Page 644-645; 2188pp; English.  
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC S. epidermidis polypeptides (II) via the production of vectors  
 CC containing them which are used to produce hosts cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identify compounds that may be used for the  
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX SQ Sequence 1151 AA;

Query Match 42.0%; Score 2465; DB 22; Length 1151;  
 Best Local Similarity 45.4%; Pred. No. 1.6e-174;  
 Matches 524; Conservative 198; Mismatches 399; Indels 34; Gaps 14;  
 Qy 26 LPAFKKILVANGETAVRAFRALLETGAATVAIYTPREDRGSHRSFSEAVRIGTESPV 85  
 Db 5 LKQIKKLLVANGETAIRIFRAAAELNSTVIAIYSDNDSKSLHRYKKADESIVLSGDLGPA 64  
 Qy 86 KAYLDEIIEIIGAARKVKADATYPGYFLSENAQLARECAENGITFIGTPPEVLDLTGDKS 145  
 Db 65 ESYLNIERIIEVALRAGVDAIHPGYFLSENEQFARRCAEEGKIFIGPHLEHLMDFGDKV 124  
 Qy 146 RAVTAAKRAKGLPVL-AESTPSKNIDDIVKSAEGQTYPIFYKAVAGGGGRMRFSVSSDEL 204  
 Db 125 KARTTAIINANLPVIGTDPGIESFEAAEQFANEAGYPLMIKATSGGGKGMRIRESSEL 184  
 Qy 205 RKLATEASKEAAAFGCGSVYVERAVINPOHIEVQILGDRTEVVVHLYERDCSLQRHOK 264  
 Db 185 EDAPHRAKSEAKSGNSEVVIERYIDNPKHIEVQVIGDEFGNIIHLHYERDCSVQRHOK 244  
 Qy 265 VVEIAPAOHLDELDRICADAVKFCISGYOGAGTVEFLV--DEKGNHVTIENNPRTQV 322  
 Db 245 VVEVAPSVCLSNKLREICDAAIQLMENIKYVNAVGVLEFLVSGDE---FFLEVNPRVQV 301  
 Qy 323 EHTVTEEVTEVDLVKAQMLAAGATL---KELGLTQDKIKTHGAALQCRITTEPDNNGFR 379  
 Db 302 EHTITMITGIDIVKTQILVADGESLFGDKISMPQONEIQTLGYAIOCRITTEPDNDFM 361  
 Qy 380 PDTGTITAYRSPGGAGVRLD-GAAQLGGEITAHFDSMLVMKTCRGSDFETAVARAQALA 438  
 Db 362 PDSGTIIAYRSRSGGFGVRLDAGDGFQGAIEISPYDLSLVKLSTHAVSFQKAEKMERSLR 421  
 Qy 439 EFTVSGVATNIGFLRALLREEDFTSKRIATGFIGDHPHLLQAPPADDEQGRILDYLDVDT 498  
 Db 422 EMRIRGVKTNIPFLINVRNDKFRSGDYTTKFIETPELFDIAPTLDRGTKTLEYIGNVT 481  
 Qy 499 VNKPHGV--RPKDVAAPIDKLPNIKDLPLPR--GSRDLKQLGPAAAFARDLREQDALAVT 554  
 Db 482 INGFNVKRPKP-EYESTKIPKISQKKINQLFGTKLEQHGPTVNTNWRREQEDVLIT 540  
 Qy 555 DTFPRDAHQSLLATRVRSFALKPAEAAVAKLTPELLSVEAMGATYDVAMRFLFEDPDWR 614  
 Db 541 DTFPRDAHQSLLATRVRTKDMNIAKTAEVFKDSFSLMWWGGATFDVAVNFKENPWR 600  
 Qy 615 LDELREAMPNVIOMLLRGRTVGYTPYDSCVAFYKAAASSGVDFRIFEDALNDVDSOM 674  
 Db 601 LERLKAIPNLVLFQMLLRASNAVYKYPDNVKKFVHESAKAGVDVFRIFDSLNWDDM 660  
 Qy 675 RPAIDAVLETNATAEAVAMAYSGD-LSDPNEKLYTLDYLLKMAEEIVKSGAHIAIKDMA 733  
 Db 661 KVANAQVEAG-MVSEGTICYTGDIILNAERSNIYTDYVVKMAKELEREGRPHILAIKDMA 719  
 Qy 734 GLLRFAAVTKLVTALRRREFDLFVHVHTHTDAGGOLATYFAAAQAGADAVDGASAPLSGTT 793  
 Db 720 GLLKPKAAAYELIGELREATHLPILHLTHDTSNGLLTYKQADAGVDLIDTAVASMSGLT 779  
 Qy 794 SQPSLSAIVAAFAHTRDRTGLSEAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIP 853

Db 780 SOPSANSLYALNGPRLNRDIDGLELSHYWSVVRPYIADFESDIKSPNTEIYQHEMP 839  
 QY 854 GGQSLNLRQAATAGLADREFELIEDNYAAVNEMLGRPTKVTSSKVVGDALHLVAGVD 913  
 Db 840 GGQYSLNLSQAQSLGLGERFEDVKEMRYRNVFLGDLVKVTPSSKVVGDALHLYMVDND 899  
 QY 914 PADFAADPKYDIPDSVIAFLRGELGNPPGGWPEPLRTRALGRSEGRAPLTPVPEEQA 973  
 Db 900 EDTVINDGVKLDPEVSVFFKGDIGPVNGFNKKLDVILKGOQ----PITERGEYLE 955  
 QY 974 HLDAD-----DSKERNNSLNRLFPKPTBEFLEHRRFGNTSALDREFFYGLV 1022  
 Db 956 PVDFAIRQELSDIOQDEVTEQDIISYVLYPKVYQYIQTKQFQGNVSLDPTFLFGMR 1015  
 QY 1023 EGRETLRLPDRVTRTLVRLDLAISPPDDKGMENVANVNGQIRPMRVRDRSVESYTAAE 1082  
 Db 1016 NGETVEIEI-DTGKRLIILKETISEPDENGKRTIYYAMNQARRIYIODENKVTNANVKP 1074  
 QY 1083 KADSSNKGHVAAFPAGVVT-VTVAGDEVKAGDAVAIAEAMKWEATITASVDGKIERRVV 1141  
 Db 1075 KADKSNPNHIGAQMGPSVTEKVSVDGDEVOANQPLLIETAMKMETIQAPDGLIKQIN 1134  
 QY 1142 PAATKVEGDLIVV 1156  
 Db 1135 ANGDAIATGDLLEVEI 1149

RESULT 15

ABP38583  
 ID ABP38583 standard; Protein; 1154 AA.  
 AC ABP38583;  
 XX  
 XX  
 DT 24-JUL-2002 (first entry)  
 XX  
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3428.  
 XX  
 KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN US6380370-B1.  
 XX  
 PD 30-APR-2002.  
 XX  
 PF 13-AUG-1998; 98US-0134001.  
 XX  
 PR 14-AUG-1997; 97US-055779P.  
 PR 08-NOV-1997; 97US-064964P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Doucette-Stamm LA, Bush D;  
 XX  
 DR WPI; 2002-381255/41.  
 DR N-PSDB; ABN91128.  
 XX  
 PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
 PT polypeptide, useful for diagnosing and treating bacterial infections -  
 XX  
 XX Disclosure; SEQ ID 3428; 267pp; English.  
 XX  
 CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life  
 CC cycle or inhibit S. epidermidis infection.  
 CC N.B. The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.  
 XX  
 SQ Sequence 1154 AA;  
 Query Match 42.0%; Score 2465; DB 23; Length 1154;  
 Best Local Similarity 45.4%; Pred. No. 1.6e-174;  
 Matches 524; Conservative 198; Mismatches 399; Indels 34; Gaps 14;  
 QY 26 LPAPKILVANGETAVAFRAAETGAATVAIYIPREDRGSHRFSASVAVRIGTEGSPV 85  
 Db 6 LKQIKLLVANGETAIRIFRAAELNISTVAIYNSNEDKSLHRYKADESILVGLDGLPA 65  
 QY 86 KAYLDIDEIIGAAGKVKADAIYVGYGFLSENAQLARECAENGITFTGTPPEVLDLTGDKS 145  
 Db 66 ESYLNERIIEIIVALAGVDIAHPGYGFLSENEQFARCAEBGKIFIGPHLEHLDMPGDKV 125  
 QY 146 RAVTAAKKAGLPV-AESTPSKNIDDIKVSAGQYPIYFVAVAGGGGRMFVSSPDEL 204  
 Db 126 KARTTAINANLPVPGTDGPIESFEAAEQFANEAGVPLMIKATSGGGGKMRIVRESSEL 185  
 QY 205 RKLATEASREAAFGDSVYVERAVINPOHIEVQLIGDRTGEVHLYERDCSLORRHK 264  
 Db 186 EDAPFRKSEAEKSEFNGSEVYERYIDNPKHIEVQVIGDEFGNIIHLYERDCSVQRHOK 245  
 QY 265 VVEIAPAOHLDPEDLRICADAVKFCRSIGYOGAGTVFVLV--DEKGNHVFTEMPIRQV 322  
 Db 246 VVEVAPSVGLSKNLRERICDAIQLMENIKYVAGTVFVLSGDE----FFFEVNPVQV 302  
 QY 323 EHTVTEVEVDLVKAQMLAAAGATL--KELGLTQDKIKTHGAALQCRITTEDPNNGPR 379  
 Db 303 EHTITEMITGIDIVKTQILVANGESLFGDKISMPQONEIQTILGYAIQCRITTEDPTNDEM 362  
 QY 380 PDTGTITAYRSGGAGVRLD-GAAQLGGEITAHFDSMLVKMTCRGSDFETAVARAQALA 438  
 Db 363 PDSGTITAYRSGGAGVRLD-GAAQLGGEITAHFDSMLVKMTCRGSDFETAVARAQALA 422  
 QY 439 EFTVSGVATNIGFRLALLREEDFTSKRIATFGIDHPHLLQAPPADDEQGRILDYLDVY 498  
 Db 423 EMRIGKVTNIPFLNVNRNDFRSQDYTKFIEETPELFDIAPTLDRGTKLEYIGNVT 482  
 QY 499 VNKPHGV--RPKDVAAPIDKLNKIDPLPR--GSRDLKQLGPAFAFARDLREQDALAVT 554  
 Db 483 INGFPNVEKRKP-EVESTKIPKISQKINQLFGTKILQEQHGPTGVTNWRREQEDVLIT 541  
 QY 555 DTFRDAHQSLLATVRSFALKPAAEAVALTPPELLSVEAWGATYDVAMRFLFEDPWDR 614  
 Db 542 DTFRDAHQSLLATVRSFALKPAAEAVALTPPELLSVEAWGATYDVAMRFLFEDPWDR 601  
 QY 615 LDELREAMPNVNIOMLLRGRTVGYTPYDPSVCRFAFVKEAASGVDIFRIFDALNDVDSOM 674  
 Db 602 LERLKAIPNVLFQMLLRASNAVGYKNYDNDVIKFFVESAKAGVDVFRIFSLNWDOM 661  
 QY 675 RPAIDAVLETNTAVAEAMAYSGD-LSDPNEKLYTLDYLYLKMAEEIVKSGAHLAIKDMA 733  
 Db 662 KVANEAVQEG-MVSEGTICYTGDILNACRSNIYLDYVYVMAKELEREGFHILAIKDMA 720  
 QY 734 GLLRPAANTKLVTLARREFDLPVHVHTHTAGCOLATYFAAAQAGADAVDGASAPLSGTT 793  
 Db 721 GLLRPAANTKLVTLARREFDLPVHVHTHTAGCOLATYFAAAQAGADAVDGASAPLSGTT 780  
 QY 794 SOPSLSAIVAAFAHTRRDTGLSEAVSDLEPVYAVRGLYLPESGTCPTGRVYRHEIP 853  
 Db 781 SOPSANSLYALNGPRLNRDIDGLELSHYWSVVRPYIADFESDIKSPNTEIYQHEMP 840  
 QY 854 GGQSLNLRQAATAGLADREFELIEDNYAAVNEMLGRPTKVTSSKVVGDALHLVAGVD 913  
 Db 841 GGQYSLNLSQAQSLGLGERFEDVKEMRYRNVFLGDLVKVTPSSKVVGDALHLYMVDND 900  
 QY 914 PADFAADPKYDIPDSVIAFLRGELGNPPGGWPEPLRTRALGRSEGRAPLTPVPEEQA 973  
 Db 901 EDTVINDGVKLDPEVSVFFKGDIGPVNGFNKKLDVILKGOQ----PITERGEYLE 956



GenCore version 5.1.4\_p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 18:59:26 ; Search time 17 seconds  
(without alignments)  
2002.489 Million cell updates/sec

Title: US-09-974-973-2

Perfect score: 5865

Sequence: 1 MTALTIGLLKGIITLVST.....RVVPAATKVEGGDLIVVVS 1157

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/ptodata/1/iaa/PTCUS\_COMB.pep.\*

6: /cgn2\_6/ptodata/1/iaa/backfiles.pep.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5759	98.2	1140	4	US-09-220-081-2
2	5759	98.2	1140	4	US-09-677-575-2
3	2465	42.0	1154	4	US-09-134-001C-3428
4	1192	20.3	694	4	US-09-433-043B-126
5	1039.5	17.7	593	4	US-09-433-043B-122
6	1032.5	17.6	447	1	US-08-611-107-6
7	1032.5	17.6	447	2	US-08-422-560A-6
8	1032.5	17.6	447	4	US-08-468-793-6
9	1005.5	17.1	453	2	US-08-611-107-8
10	1005.5	17.1	453	2	US-08-422-560A-8
11	1005.5	17.1	453	4	US-08-468-793-8
12	1003.5	17.1	453	4	US-09-433-043B-121
13	996.5	17.0	453	1	US-07-956-700B-6
14	996.5	17.0	453	1	US-08-476-537-6
15	996.5	17.0	453	1	US-08-485-607-6
16	996.5	17.0	453	2	US-08-475-879-6
17	996.5	17.0	453	4	US-09-433-043B-6
18	954	16.3	605	4	US-09-433-043B-123
19	940	16.0	448	1	US-08-074-121-3
20	940	16.0	448	5	PCT-US94-06447-3
21	933	15.9	427	1	US-07-956-700B-3
22	933	15.9	427	1	US-08-476-537-3
23	933	15.9	427	1	US-08-485-607-3
24	933	15.9	427	2	US-08-475-879-3
25	933	15.9	427	4	US-09-433-043B-3
26	928	15.8	536	4	US-08-662-344-2
27	921	15.7	449	1	US-08-074-121-6

28	921	15.7	449	5	PCT-US94-06447-6
29	849	14.5	461	4	US-09-134-001C-3604
30	842.5	14.4	722	4	US-09-433-043B-125
31	786	13.4	703	4	US-09-433-043B-124
32	782	13.3	411	4	US-09-134-001C-3221
33	613	10.5	853	4	US-09-433-043B-119
34	605	10.3	852	4	US-09-433-043B-118
35	583.5	9.9	2257	1	US-08-611-107-10
36	583.5	9.9	2257	2	US-08-422-560A-10
37	583.5	9.9	2257	4	US-08-468-793-10
38	561.5	9.6	2325	3	US-08-417-089-6
39	561.5	9.6	2325	4	US-08-695-651-6
40	561.5	9.6	2325	4	US-08-930-285-6
41	561.5	9.6	2325	4	US-08-695-421-6
42	561.5	9.6	2325	4	US-08-697-826A-10
43	559	9.5	2172	1	US-08-611-107-31
44	555.5	9.5	2254	2	US-08-677-010-3
45	555.5	9.5	2254	2	US-08-790-519-3

ALIGNMENTS

RESULT 1

US-09-220-081-2  
; Sequence 2, Application US/09220081  
; Patent No. 6171833  
; GENERAL INFORMATION:  
; APPLICANT: Sinskey, Anthony J.  
; APPLICANT: Lessard, Philip A.  
; APPLICANT: Willis, Laura B.  
; APPLICANT: Stephanopoulos, Gregory  
; TITLE OF INVENTION: Pyruvate Carboxylase from Corynebacterium glutamicum  
; FILE REFERENCE: 1533.0790000  
; CURRENT APPLICATION NUMBER: US/09/220,081  
; CURRENT FILING DATE: 1998-12-23  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 1140  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-220-081-2

Query Match	98.2%	Score	5759	DB	4	Length	1140
Best Local Similarity	99.4%	Pred. No.	0				
Matches	1133	Conservative	5	Mismatches	2	Indels	0
Gaps	0						
Qy	18	VSTHTSSTLPAPFKKILVANRGEIAVRAALETGAATVAIYPREDRGSFHRSEAVR	77				
Db	1	MSTHTSSTLPAPFKKILVANRGEIAVRAALETGAATVAIYPREDRGSFHRSEAVR	60				
Qy	78	IGTEGSPKAYLDIDEITGAARKVKADAIYPGYFLSNAQLARECAENGITFGTPEV	137				
Db	61	ICTEGSPKAYLDIDEITGAARKVKADAIYPGYFLSNAQLARECAENGITFGTPEV	120				
Qy	138	LDLTGDKSRVTAATAKAGLPVLAESTPSKNIDDIIVKSEAGOTYPIFVKAVAGGGGRMRF	197				
Db	121	LDLTGDKSRVTAATAKAGLPVLAESTPSKNIDDIIVKSEAGOTYPIFVKAVAGGGGRMRF	180				
Qy	198	VSSPDRLKRLATEASREAAAFDGGSVYVERAVINPQIHVEQVILGDRGTGEVHLYERDCS	257				
Db	181	VASPDRLKRLATEASREAAAFDGGSVYVERAVINPQIHVEQVILGDRGTGEVHLYERDCS	240				
Qy	258	LQRHQQVVEIAPQHLDELDRICADAVKFCRSIGYQAGTVEFLVDEKGNHVFIEMN	317				
Db	241	LQRHQQVVEIAPQHLDELDRICADAVKFCRSIGYQAGTVEFLVDEKGNHVFIEMN	300				
Qy	318	PRIQVEHVTVEEVDLVKQMLAAGATLKLGLTODKIKTHGAALQCRITTEDPNNG	377				
Db	301	PRIQVEHVTVEEVDLVKQMLAAGATLKLGLTODKIKTHGAALQCRITTEDPNNG	360				
Qy	378	FRPDTGTTITVRSPPGAGVRLDGAALQGLGEITAHFDSMLVKMTCRGSDFETAVARAQAL	437				

Db 361 FRPDTGTTAYRSPGGAGVRLDGAALGGGTAHFDMSLVKMTKCGSDFTAVARAQAL 420  
QY 438 AEFVSVATNIGFLRALLREEDFTSKRIATGFIQDHPHLLQAPPADEQGRILLYLADV 497  
Db 421 AEFVSVATNIGFLRALLREEDFTSKRIATGFIADHPHLLQAPPADEQGRILLYLADV 480  
QY 498 TVNPKHGVREPKDVAAPDKLPNTKDLPLPRGSRDLKQLGPAAFARDLRQDALAVTDFT 557  
Db 481 TVNPKHGVREPKDVAAPDKLPNTKDLPLPRGSRDLKQLGPAAFARDLRQDALAVTDFT 540  
QY 558 FRDAHQSLIATRVRSFALPAAAVAKLIPPELLSVFAGWGATYDVAMRFLFEDPDWRDLDE 617  
Db 541 FRDAHQSLIATRVRSFALPAAAVAKLIPPELLSVFAGWGATYDVAMRFLFEDPDWRDLDE 600  
QY 618 LREAMPNVNTOMLLRGRNTVGYTYPDSCVCRAPVKEAASGVVDIFRIFDALNDVSQMRPA 677  
Db 601 LREAMPNVNTOMLLRGRNTVGYTYPDSCVCRAPVKEAASGVVDIFRIFDALNDVSQMRPA 660  
QY 678 IDAVLETTNTAVAEVAMAYSGDLSDPNEKLYTLDYLLKMAEEIVKSGAHILAIKDMAGLLR 737  
Db 661 IDAVLETTNTAVAEVAMAYSGDLSDPNEKLYTLDYLLKMAEEIVKSGAHILAIKDMAGLLR 720  
QY 738 PAAVTKLVTAALRRREFDLPVHVHTDHTAGGOLATYFAAAQAGADAVDCASAPLSGTTSQPS 797  
Db 721 PAAVTKLVTAALRRREFDLPVHVHTDHTAGGOLATYFAAAQAGADAVDCASAPLSGTTSQPS 780  
QY 798 LSAIVAAFAHTRDTGLSLSEAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGGOL 857  
Db 781 LSAIVAAFAHTRDTGLSLSEAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGGOL 840  
QY 858 SNLRQAQATLGLADREFELIEDNYAANVEMLGRPTKVTPESSKVVGDGLALHLVAGVDPADF 917  
Db 841 SNLRQAQATLGLADREFELIEDNYAANVEMLGRPTKVTPESSKVVGDGLALHLVAGVDPADF 900  
QY 918 AADPOKYDIPDSVIAFLRGELNPPGWPPELRTALRGRSEKAPLTPVPEEQAHLLA 977  
Db 901 AADPOKYDIPDSVIAFLRGELNPPGWPPELRTALRGRSEKAPLTPVPEEQAHLLA 960  
QY 978 DSKERRNSLNRLLFPKPTPEEFLEHRRRNGTSALDDREFFYGLVGEGRETLIRLPDVRTP 1037  
Db 961 DSKERRNSLNRLLFPKPTPEEFLEHRRRNGTSALDDREFFYGLVGEGRETLIRLPDVRTP 1020  
QY 1038 LLVRLDAISEPDDKGRNVVANYNGQIRPMRVDRDSVESVTATAEKADSSNKGHVAAPFA 1097  
Db 1021 LLVRLDAISEPDDKGRNVVANYNGQIRPMRVDRDSVESVTATAEKADSSNKGHVAAPFA 1080  
QY 1098 GVVTVTVAEGDEVKAGDAVAIEAMKMEATITASVDGKIERYVVVPAATKVEGGDLIVVVS 1157  
Db 1081 GVVTVTVAEGDEVKAGDAVAIEAMKMEATITASVDGKIERYVVVPAATKVEGGDLIVVVS 1140

## RESULT 2

US-09-677-575-2  
: Sequence 2, Application US/09677575  
: Patent No. 6403351  
: GENERAL INFORMATION:  
: APPLICANT: Sinskey, Anthony J.  
: APPLICANT: Lessard, Philip A.  
: APPLICANT: Willis, Laura B.  
: APPLICANT: Stephanopoulos, Gregory  
: TITLE OF INVENTION: Pyruvate Carboxylase from *Corynebacterium glutamicum*  
: FILE REFERENCE: 1533.0790000  
: CURRENT APPLICATION NUMBER: US/09/677,575  
: CURRENT FILING DATE: 2000-10-03  
: PRIOR FILING DATE: 1998-12-23  
: NUMBER OF SEQ ID NOS: 36  
: SOFTWARE: PatentIn Ver. 2.0  
: SEQ ID NO 2  
: LENGTH: 1140  
: TYPE: prt  
: ORGANISM: *Corynebacterium glutamicum*

US-09-677-575-2

Query Match 98.2%; Score 5759; DB 4; Length 1140;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1133; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
QY 18 VSTHTSSLTLPFAFKTLVANRGEIAVRAFAALETGAATVATVYIPREDRSGSPHRSFASAVR 77  
Db 1 MSTHTSSLTLPFAFKTLVANRGEIAVRAFAALETGAATVATVYIPREDRSGSPHRSFASAVR 60  
QY 78 IGTEGSPVKAYLDIDEIIGAARKKADAIYCYGFLSENAQLARECAENGITFTGPTPEV 137  
Db 61 IGTEGSPVKAYLDIDEIIGAARKKADAIYCYGFLSENAQLARECAENGITFTGPTPEV 120  
QY 138 LDLTGDKSRVTAARKAGLPVLAESTPSKNIIDIVKSAEGQYPIFYKAVAGGGGMRMF 197  
Db 121 LDLTGDKSRVTAARKAGLPVLAESTPSKNIIDIVKSAEGQYPIFYKAVAGGGGMRMF 180  
QY 198 VSSPDELKRLATEASREAAAFGDSGVVVERAVINPQIHIEVQIILGDRTGEVHVLYERDCS 257  
Db 181 VASPDDELKRLATEASREAAAFGDSGVVVERAVINPQIHIEVQIILGDRTGEVHVLYERDCS 240  
QY 258 LQRHQKVVLEIAPAQHLDPELRDRICADAVKFCRSIGYQGAGTVEFLVDEKGNHVFTEMN 317  
Db 241 LQRHQKVVLEIAPAQHLDPELRDRICADAVKFCRSIGYQGAGTVEFLVDEKGNHVFTEMN 300  
QY 318 PRIOVEHTVTEVEVDLVKQMRLAAGATLKLGLTQDKIKTHGAALOCRTITTEDPNNG 377  
Db 301 PRIOVEHTVTEVEVDLVKQMRLAAGATLKLGLTQDKIKTHGAALOCRTITTEDPNNG 360  
QY 378 PRPDGTITAYRSPGGAGVRLDGAALGGGTAHFDMSLVKMTKCGSDFTAVARAQAL 437  
Db 361 PRPDGTITAYRSPGGAGVRLDGAALGGGTAHFDMSLVKMTKCGSDFTAVARAQAL 420  
QY 438 AEFVSVATNIGFLRALLREEDFTSKRIATGFIQDHPHLLQAPPADEQGRILLYLADV 497  
Db 421 AEFVSVATNIGFLRALLREEDFTSKRIATGFIADHPHLLQAPPADEQGRILLYLADV 480  
QY 498 TVNPKHGVREPKDVAAPDKLPNTKDLPLPRGSRDLKQLGPAAFARDLRQDALAVTDFT 557  
Db 481 TVNPKHGVREPKDVAAPDKLPNTKDLPLPRGSRDLKQLGPAAFARDLRQDALAVTDFT 540  
QY 558 FRDAHQSLIATRVRSFALPAAAVAKLIPPELLSVFAGWGATYDVAMRFLFEDPDWRDLDE 617  
Db 541 FRDAHQSLIATRVRSFALPAAAVAKLIPPELLSVFAGWGATYDVAMRFLFEDPDWRDLDE 600  
QY 618 LREAMPNVNTOMLLRGRNTVGYTYPDSCVCRAPVKEAASGVVDIFRIFDALNDVSQMRPA 677  
Db 601 LREAMPNVNTOMLLRGRNTVGYTYPDSCVCRAPVKEAASGVVDIFRIFDALNDVSQMRPA 660  
QY 678 IDAVLETTNTAVAEVAMAYSGDLSDPNEKLYTLDYLLKMAEEIVKSGAHILAIKDMAGLLR 737  
Db 661 IDAVLETTNTAVAEVAMAYSGDLSDPNEKLYTLDYLLKMAEEIVKSGAHILAIKDMAGLLR 720  
QY 738 PAAVTKLVTAALRRREFDLPVHVHTDHTAGGOLATYFAAAQAGADAVDCASAPLSGTTSQPS 797  
Db 721 PAAVTKLVTAALRRREFDLPVHVHTDHTAGGOLATYFAAAQAGADAVDCASAPLSGTTSQPS 780  
QY 798 LSAIVAAFAHTRDTGLSLSEAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGGOL 857  
Db 781 LSAIVAAFAHTRDTGLSLSEAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGGOL 840  
QY 858 SNLRQAQATLGLADREFELIEDNYAANVEMLGRPTKVTPESSKVVGDGLALHLVAGVDPADF 917  
Db 841 SNLRQAQATLGLADREFELIEDNYAANVEMLGRPTKVTPESSKVVGDGLALHLVAGVDPADF 900  
QY 918 AADPOKYDIPDSVIAFLRGELNPPGWPPELRTALRGRSEKAPLTPVPEEQAHLLA 977  
Db 901 AADPOKYDIPDSVIAFLRGELNPPGWPPELRTALRGRSEKAPLTPVPEEQAHLLA 960  
QY 978 DSKERRNSLNRLLFPKPTPEEFLEHRRRNGTSALDDREFFYGLVGEGRETLIRLPDVRTP 1037  
Db 961 DSKERRNSLNRLLFPKPTPEEFLEHRRRNGTSALDDREFFYGLVGEGRETLIRLPDVRTP 1020

Query Match 20.3%; Score 1192; DB 4; Length 694;  
Best Local Similarity 27.9%; Pred. No. 6.6e-86;  
Matches 316; Conservative 102; Mismatches 239; Indels 476; Gaps 11;

```
QY 31 KILVANGETIAVRAFALETAATVAIYPREDSGPHRSFASAEVRICTEG--SPVKAY 88
DB 21 KILVANGETIPIRIFRTAHELMSQTAIYSHEDRLSTHKKADAEVYIGVGQYTPGAY 80
QY 89 LDIDEITGAKKVADAIYPGYGLSENQAOLARECAENGITFIPTPEVLDLTGDKSRV 148
DB 81 LAIDEITIAQKHQVDTHPGYGLSENSEFADKVKAGITWIGPAPAEVDSVGDVKSR 140
QY 149 TAAKAGLPVLAEST-PSKNIDDIVKSAEGOTYPIFYKAVAGGGGMRVSPDELKRL 207
DB 141 NLAAKANVTYVPGPIETVEEALDFVNEYGYPIVIAKAAFGGGGMRVREGDDVADA 200
QY 208 ATEASREAEAFSGSVYVVERAVINPOHIEVQILGDRTEGVVHLYERDCSLORRHQVVE 267
DB 201 FORATSEARTAGFNGTCFEVERFDKPKHIEVQLLADNHNHNVHLFERDCSVQRHOKVVE 260
QY 268 IAPAQHLDPELDRICADAVKFCRSIGYOGAGTVVEFLVDEKGNHVFTEMNPRIOVEHTV 327
DB 261 VAPAKTLPREVDRDAILTDVAKLAKECYRNAGTAELFLVDNQRHVFIEINPRIOVEHTIT 320
QY 328 EEVTEVDLVKAOMLAAGATLKEGLTQDKIKTHGAALQCRITTEDPNNNGFRPDTGTTA 387
DB 321 EETGIDIVAAQIQ-AAGASLPQLGLFQDKITTRGFALQCRITTEDPAKNFQPDGTGRIEV 379
QY 388 YRSPGGAGVLDGA-AOLGEITAHFDSMLVKMTCRGSDFETAVARAORALAEFTVSGVA 446
DB 380 YRSAGGNVRLDGNAYAGTIISPHYDSMLVKSCSGSYEIVRRKMIRALIEFRIGVK 439
QY 447 TNIGFLRALLREEDFTSKRTATGFIGDHPHLLQAPPADDEQGRILDYLDVTVNKKPHGVR 506
DB 440 TNIFELTLTNVFIETGVTGTFIDTQPOLFOVNSONRAQLLHLYADVA----- 491
QY 507 PKOVAAPIDKLPNIKDLPLRGSRDRLKQLGPAFAFARDLREQDALAVTDTFFRDAHOSLL 566
DB 492 ----- 491
QY 567 ATRVRSFALKPAEAVAKLTPELLSVBANGATYDVAMRFLFEDPMDRLDELREAMPNVN 626
DB 492 ----- 491
QY 627 IQMLLRGRNTVGYTYPDSVCRAPFVKEAASSGVDFIRFDALNDVDSQMRPAIDAVLETNT 686
DB 492 ----- 491
QY 687 AVEAVAMAYSGDLSDPNEKLYTLDYLLKMAEIVKSGAHLIAIKDNAGLLRPAAVTKLVT 746
DB 492 ----- 491
QY 747 ALRREFDLPVHVHTHTAGQLATYFAAQAGADAVDGSAPLSGTTSQPSLSAIVAAFA 806
DB 492 ----- 491
QY 807 HTRDRTGLSLEAVSDLEPYEAVRGLYLPFESGTPGTGRVYRHETPGGQLSMLRAQATA 866
DB 492 ----- 491
QY 867 LGLADREFLEDNYAANEMLGRP-TKVTPSSKVVGDALHLVAGVDPADFAADPOKYD 925
DB 492 ----- 491
QY 926 IPDSVIAFLRGELGNPPGPEPLRTRALEGRSEKAPLTVPEEPEQAHLDDADSKERN 985
DB 518 ----- 517
QY 986 SLNRLLPKPTPEFLEHRRRRFGNTSALDDREFFYGLVEGRETLIRLPDVRTPLLVRLDAI 1045
DB 518 ----- 517
QY 1046 SEPDDK-GMKNVYANVGOILRPMRVDRSVESVTATAEKADSNKGNHVAAPFAGV-VTVT 1103
DB 572 GDLNKKTGEREVFDLNGEMRKTRVADRSQKVETVTKSRADMHDPLHIGAPMAGVIVEVK 631
```

RESULT 6

```
QY 1104 VAEGDEVKAGDAVAIIEMAKMEATITASVDGKIERYVVPAAATKVEGGDLIVVV 1156
DB 632 VHKGSLITKGGQPVAVLSAMKMEMIISPSDQGVKEVFSVDGENVDSSOLLVLL 684

RESULT 5
US-09-433-043B-122
; Sequence 122, Application US/09433043B
; Patent No. 6399342
; GENERAL INFORMATION:
; APPLICANT: HASELKORN, ROBERT
; TITLE OF INVENTION: CYANOBACTERIAL AND PLANT ACETYL-CoA CARBOXYLASE
; FILE REFERENCE: ARCD:338US
; CURRENT APPLICATION NUMBER: US/09/433,043B
; CURRENT FILING DATE: 1999-10-25
; PRIOR APPLICATION NUMBER: 08/475,879
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 07/956,700
; PRIOR FILING DATE: 1992-10-02
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 122
; LENGTH: 593
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Peptide
US-09-433-043B-122

Query Match 17.7%; Score 1039.5; DB 4; Length 593;
Best Local Similarity 44.3%; Pred. No. 6.3e-74;
Matches 229; Conservative 80; Mismatches 177; Indels 31; Gaps 6;

QY 29 FKILVANGETIAVRAFALETAATVAIYPREDSGPHRSFASAEVRICTEGSPVKAY 88
DB 3 FDKILIANGETIALRLRACEEMGIATIAVSTVDNRNALHVQLADEAVCIG-EPASAKSY 61
QY 89 LDIDEITGAKKVADAIYPGYGLSENQAOLARECAENGITFIPTPEVLDLTGDKSRV 148
DB 62 LNIPNIAAALTRNASAIHPGYGLSENKAEICADHIAFIPTPEAIRLMGDKSTAK 121
QY 149 TAAKAGLPVLAES-----TPSKNIDDIVKSAEGOTYPIFYKAVAGGGGMRVFSVSPDE 203
DB 122 ETMQAGVPTVPGSEGLVETEQGLE---LAKDIGYPMIKATAGGGRMRLVRSPE 177
QY 204 LRKLATEASREAEAFSGSVYVVERAVINPOHIEVQILGDRTEGVVHLYERDCSLORRH 263
DB 178 FVKFLAQAGEAGAFAGNAGYTIERPRHIEFQIILADNVGNVHILGERDCSTQRNQ 237
QY 264 KVVETAPAOHLDPELDRICADAVKFCRSIGYOGAGTVVEFLVDEKGNHVFTEMNPRIOVE 323
DB 238 KLEESAPSPALDSDLREKMGQAAVRAAQAFINYAGAGTIEFLDRSGQYFMEMNTRIOVE 297
QY 324 HVTVEVTEVDLVKAOMRLAAGATLKEGLTQDKIKTHGAALQCRITTEDPNNNGFRPDTG 383
DB 298 HPTVETMTGVDLLVEQIRIAQGERLR--LTQDQYVLRGHAIECRINAEDPDHDFRPAPG 354
QY 384 TITAYRSPGGAGVRLDGAALGGEITAHFDSMLVKMTCRGSDFETAVARAORALAEFTVS 443
DB 355 RISGGLPFGGVRIDSHVYTDYQIPPYDSLIGKLVWGPDRATAINRMKRALRECAIT 414
QY 444 GVATNTGFLRLREDEFTSKRIATGFIGDHPHLLQAPPADDEQGRILDYLA----- 495
DB 415 GLPTTIGFHORIMENPQFLOGNVSTSFVOE---MNKPLDFNEIRQLLTIAQTADIAEVT 470
QY 496 -----DVTNKPCHVRPKDVAAPIDKLPNKKDPLP 526
DB 471 LKSDDFELTVKAVGNVNSVVPVVTAPLSGVVSGCLP 507

RESULT 6
```



QY	324	HTVTEEVTDVLYKAQMR	LAAGATLKLGLTQDKIKTHGAALQCRITTEDPNNNGFRPDYG	381
Db	298	HPVTVMVTGVDLLVEQIRIA	GERLR---LTQDVVLVGRHAIECRINAEDPDHDFRPAPG	354
QY	384	TITAYRSPGAGVRLDGAAL	GGGTTAHTFDSMLVWTCRSDFTAVARAQALAEFTVS	443
Db	355	RISGYLPPGGVGRIDSHV	YTYDQIPYDLSGLKLVWGPDRATINRMKRALRECAIT	414
QY	444	GVATNIGFLRALLREEDFT	SKRIATGFI	471
Db	415	GLPTTIGFHQRIMENPQ	FLOGNVSTSEV	442
RESULT 7				
US-08-422-560A-6				
; Sequence 6, Application US/08422560A				
; Patent No. 5910626				
; GENERAL INFORMATION:				
; APPLICANT: Haselkorn, Robert				
; APPLICANT: Gornicki, Piotr				
; TITLE OF INVENTION: ACETYL-CoA CARBOXYLASE COMPOSITIONS AND				
; TITLE OF INVENTION: METHODS FOR USE				
; NUMBER OF SEQUENCES: 31				
; CORRESPONDENCE ADDRESS:				
; ADDRESSEE: Arnold, White & Durkee				
; STREET: P.O. Box 4433				
; CITY: Houston				
; STATE: TX				
; COUNTRY: USA				
; ZIP: 77210-4433				
; COMPUTER READABLE FORM:				
; MEDIUM TYPE: Floppy disk				
; COMPUTER: IBM PC compatible				
; OPERATING SYSTEM: PC-DOS/MS-DOS				
; SOFTWARE: PatentIn Release #1.0, Version #1.30				
; CURRENT APPLICATION DATA:				
; APPLICATION NUMBER: US/08/422,560A				
; FILING DATE: 14-APR-1995				
; CLASSIFICATION: 800				
; PRIOR APPLICATION DATA:				
; APPLICATION NUMBER: US 07/956,700				
; FILING DATE: 02-OCT-1992				
; ATTORNEY/AGENT INFORMATION:				
; NAME: Wilson, Mark B.				
; REGISTRATION NUMBER: 37,259				
; REFERENCE/DOCKET NUMBER: ARCD:152/WIM				
; TELECOMMUNICATION INFORMATION:				
; TELEPHONE: 512-418-3000				
; TELEFAX: 512-474-7577				
; INFORMATION FOR SEQ ID NO: 6:				
; SEQUENCE CHARACTERISTICS:				
; LENGTH: 447 amino acids				
; TYPE: amino acid				
; STRANDEDNESS:				
; TOPOLOGY: linear				
US-08-611-107-6				
Query Match 17.6%; Score 1032.5; DB 1; Length 447;				
Best Local Similarity 48.2%; Pred. No. 1.4e-73;				
Matches 216; Conservative 71; Mismatches 148; Indels 13; Gaps 4;				
QY	29	FKKILVANRGEIAVRAFAA	LETGAATVAIYPREDRGSFHRSEAFSEAVRIGTEGSPVKAY	88
Db	3	FDKILIANRGEIALRILRA	CEEMGIATIAVHSTVDNRNALHVQLADEAVCIQ-EPASAKSY	61
QY	89	LDIDEITGAAKVKADAI	PCGYFLSENAQLARECAENGITFIGPTPEVLDTGDKSRV	148
Db	62	LNIPNIIAALTRNAAI	HPGYGFLSENAKFAECADHIIAFIGTTPAIRLMDGKSTAK	121
QY	149	TAACKAGLPLVAES----	TPSKNIDDIVKSAEGQTYPIFKVAVAGGGGRMRFFVSSPDE	203
Db	122	ETMQAGVPTVPGSEGL	VETEQEGLE---LAKDIGYPMVKATAGGGGRMRLVRSDE</	

```

QY 204 LRKLTASREAEAAFGDGSVYVERAVINPOHIEVQILGDRGTGEVHLVYERDCSLORRHQ 263
DB 178 FVKFLAAGGAGAAFGAGVYIEKTERPRHIEFQILADNYGNVHLGRDCSIORRNQ 237
QY 264 KVVETAPQHLDPRLDRICADAVKFCRSIGYOCAGTVEFLVDEKGNHVFIEIENPRIQVE 323
DB 238 KLEEAAPSPALDSLRKMGQAQVAAQAFINYTGAGTIEFLDRSGQFYFEMENRIQVE 297
QY 324 HTVTTEVTEVDLVKAOMRLAAGATLKELGLTODKIKTHGAALOCRTITDPNNGFRPDG 383
DB 298 HPVTMTGVDLLVEQIRIAOGERLR---LTQDQVVLGRHAEICRINAEDPDHDFRPPG 354
QY 384 TITAYRSPGAGVRLDGAALGGEITAHFDSMLVKWTCRSDDETAVARAORALAEFTVS 443
DB 355 RISGLPFGPGVGRIDSHVYTDQIPYDLSLIGLKVWGPDRATINRMKRALRECAIT 414
QY 444 GVATNIGFLRALLREEDFTSKRIATGFI 471
DB 415 GLPTTIGFHQIRIMENPQFLQGNVSTSFV 442

RESULT 8
US-08-468-793-6
; Sequence 6, Application US/08468793
; Patent No. 6177267
; GENERAL INFORMATION:
; APPLICANT: Haselkorn, Robert
; APPLICANT: Gornicki, Piotr
; TITLE OF INVENTION: ACETYL-COA CARBOXYLASE COMPOSITIONS AND
; TITLE OF INVENTION: METHODS OF USE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,793
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,560
; FILING DATE: 14-APR-1995
; APPLICATION NUMBER: US SN 07/956,700
; FILING DATE: 02-OCT-1992
; CLASSIFICATION: 800
; APPLICATION NUMBER: PCT/US93/09340
; FILING DATE: 30-SEP-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: ARCD:152/KIT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-468-793-6

```

```

Query Match 17.6%; Score 1032.5; DB 4; Length 447;
Best Local Similarity 48.2%; Pred. No. 1.4e-73;
Matches 216; Conservative 71; Mismatches 148; Indels 13; Gaps 4;
QY 29 FKKILVANGETEIAVAFRAALETGAATVAIYPREDRGSFHRFSFASEAVRIGTEGSPVKAY 88
DB 3 FDKILLIANRGETALRILRACEEMGIATIAVHSTVDNRNALHVOLADEAVCIG-EPASAKSY 61
QY 89 LDIDETIGAAGKVKADATYPGVGFSENAQLAARECAENGITFPTPEVLDLTGDKSRAY 148
DB 62 LNPINIAAALFENASAHFPGVGFSENAKFAEICADHUIAFIGTTPAIRLMDKSTAK 121
QY 149 TAAKAGLPLVAES-----TPSKNIDDDIVKSAGQTYPIFVAVAGGGGRMRFFVSSPDE 203
DB 122 ETMQKAGVTVPVSGSEGLVETEQEGLE---LAKDIGVPMIKATAGGGGRMRLVRSDE 177
QY 204 LRKLTASREAEAAFGDGSVYVERAVINPOHIEVQILGDRGTGEVHLVYERDCSLORRHQ 263
DB 178 FVKFLAAGGAGAAFGAGVYIEKTERPRHIEFQILADNYGNVHLGRDCSIORRNQ 237
QY 264 KVVETAPQHLDPRLDRICADAVKFCRSIGYOCAGTVEFLVDEKGNHVFIEIENPRIQVE 323
DB 238 KLEEAAPSPALDSLRKMGQAQVAAQAFINYTGAGTIEFLDRSGQFYFEMENRIQVE 297
QY 324 HTVTTEVTEVDLVKAOMRLAAGATLKELGLTODKIKTHGAALOCRTITDPNNGFRPDG 383
DB 298 HPVTMTGVDLLVEQIRIAOGERLR---LTQDQVVLGRHAEICRINAEDPDHDFRPPG 354
QY 384 TITAYRSPGAGVRLDGAALGGEITAHFDSMLVKWTCRSDDETAVARAORALAEFTVS 443
DB 355 RISGLPFGPGVGRIDSHVYTDQIPYDLSLIGLKVWGPDRATINRMKRALRECAIT 414
QY 444 GVATNIGFLRALLREEDFTSKRIATGFI 471
DB 415 GLPTTIGFHQIRIMENPQFLQGNVSTSFV 442

RESULT 9
US-08-611-107-8
; Sequence 8, Application US/08611107
; Patent No. 5801233
; GENERAL INFORMATION:
; APPLICANT: Haselkorn, Robert
; APPLICANT: Gornicki, Piotr
; TITLE OF INVENTION: NUCLEIC ACID COMPOSITIONS ENCODING
; TITLE OF INVENTION: ACETYL-COA CARBOXYLASE AND USES
; TITLE OF INVENTION: THEREFOR
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/611,107
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US SN 07/956,700
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US93/09340
; FILING DATE: 30-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US SN 08/422,560

```

```
; FILING DATE: 14-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Kitchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: ARCD:221
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 453 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-611-107-8

Query Match 17.1%; Score 1005.5; DB 1; Length 453;
Best Local Similarity 47.1%; Pred. No. 2e-71;
Matches 209; Conservative 72; Mismatches 158; Indels 5; Gaps 3;

Qy 29 FKKILVANRGEIATVAFRAALETGAATVAIYPREDRGSFHRSEAFSEAVRIGTEGSPVKAY 88
Db 3 FKKILIANRGEIATVAFRAALETGAATVAIYPREDRGSFHRSEAFSEAVRIGTEGSPVKAY 88
Qy 89 LDIDEITGAAKKVKADAIYPGYFLSENAQLARECAENGITFPGTPEVLDLTGDKSRV 148
Db 62 LNIPIIAAALTRNASAIHPGFLAENAREFAEICADHLLTFIGSPDSIRAMGDKSTAK 121
Qy 149 TAAKAGLPVLAESTP-SKNIDDIIVKSAGQTYPIFYKAVAGGGGRGMRFFVSSPDELRLK 207
Db 122 ETMRGVGPTIPGSDGLLTVDVSAKVAEEIGYPMVKATAGGGGRGMRFLVREPADLEKL 181
Qy 208 ATEASREAAAFGDSGVYVERAVINPOHIEVQILGDRTEGVVHLYERDCSLQRHOKVVE 267
Db 182 FLAAGGEAAAFGPNGLYLEKFDPRHVEFQILADAGNVVHLCERDCSIQRHOKLLE 241
Qy 268 IAPAQHLDPELRDICAADAVKFCRSIGYGAGTVEFLVDEKGNHVFIEIENPRIQVEHTVT 327
Db 242 EAPSPALSADLRQKMGDAVVAQAIGYIGAGTVEFLVDATGTFVEMEMNTRIQVEHPT 301
Qy 328 EEVTEVDLVKAQMLAAGATLKEGLTQDKIKTHGAALQCRITTEDPNNGRPDGTGTTA 387
Db 302 EMITGLDLIAEQIRIAQGEALR---FROADIQLRGHATECRINAEDPEYFNRPNGRITG 358
Qy 388 YRSPGAGVRLDGAALGGEITAHFDSMLVKMTCRGSDFTAVARAQALAEFTVSGVAT 447
Db 359 YLPPGGVGRVDSHYVTDYEIPPYDLSLIGKLIWNGATREEARIMQALRECAITGLPT 418
Qy 448 NIGFLRALLREEDFTSKRIATGFI 471
Db 419 TLSFHQLMLQMPFELRGELYTNFV 442

RESULT 10
US-08-422-560A-8
; Sequence 8, Application US/08422560A
; Patent No. 5910626
; GENERAL INFORMATION:
; APPLICANT: Haselkorn, Robert
; APPLICANT: Gornicki, Piotr
; TITLE OF INVENTION: ACETYL-CoA CARBOXYLASE COMPOSITIONS AND
; TITLE OF INVENTION: METHODS FOR USE
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/422,560A
; FILING DATE: 14-APR-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/956,700
; FILING DATE: 02-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: ARCD:152/WIM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512-418-3000
; TELEFAX: 512-474-7577
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 453 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-422-560A-8

Query Match 17.1%; Score 1005.5; DB 2; Length 453;
Best Local Similarity 47.1%; Pred. No. 2e-71;
Matches 209; Conservative 72; Mismatches 158; Indels 5; Gaps 3;

Qy 29 FKKILVANRGEIATVAFRAALETGAATVAIYPREDRGSFHRSEAFSEAVRIGTEGSPVKAY 88
Db 3 FKKILIANRGEIATVAFRAALETGAATVAIYPREDRGSFHRSEAFSEAVRIGTEGSPVKAY 88
Qy 89 LDIDEITGAAKKVKADAIYPGYFLSENAQLARECAENGITFPGTPEVLDLTGDKSRV 148
Db 62 LNIPIIAAALTRNASAIHPGFLAENAREFAEICADHLLTFIGSPDSIRAMGDKSTAK 121
Qy 149 TAAKAGLPVLAESTP-SKNIDDIIVKSAGQTYPIFYKAVAGGGGRGMRFFVSSPDELRLK 207
Db 122 ETMRGVGPTIPGSDGLLTVDVSAKVAEEIGYPMVKATAGGGGRGMRFLVREPADLEKL 181
Qy 208 ATEASREAAAFGDSGVYVERAVINPOHIEVQILGDRTEGVVHLYERDCSLQRHOKVVE 267
Db 182 FLAAGGEAAAFGPNGLYLEKFDPRHVEFQILADAGNVVHLCERDCSIQRHOKLLE 241
Qy 268 IAPAQHLDPELRDICAADAVKFCRSIGYGAGTVEFLVDEKGNHVFIEIENPRIQVEHTVT 327
Db 242 EAPSPALSADLRQKMGDAVVAQAIGYIGAGTVEFLVDATGTFVEMEMNTRIQVEHPT 301
Qy 328 EEVTEVDLVKAQMLAAGATLKEGLTQDKIKTHGAALQCRITTEDPNNGRPDGTGTTA 387
Db 302 EMITGLDLIAEQIRIAQGEALR---FROADIQLRGHATECRINAEDPEYFNRPNGRITG 358
Qy 388 YRSPGAGVRLDGAALGGEITAHFDSMLVKMTCRGSDFTAVARAQALAEFTVSGVAT 447
Db 359 YLPPGGVGRVDSHYVTDYEIPPYDLSLIGKLIWNGATREEARIMQALRECAITGLPT 418
Qy 448 NIGFLRALLREEDFTSKRIATGFI 471
Db 419 TLSFHQLMLQMPFELRGELYTNFV 442

RESULT 11
US-08-468-793-8
; Sequence 8, Application US/08468793
; Patent No. 6177267
; GENERAL INFORMATION:
; APPLICANT: Haselkorn, Robert
; APPLICANT: Gornicki, Piotr
; TITLE OF INVENTION: ACETYL-CoA CARBOXYLASE COMPOSITIONS AND
; TITLE OF INVENTION: METHODS OF USE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
```

STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/468,793  
FILING DATE: 06-JUN-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/422,560  
FILING DATE: 14-APR-1995  
APPLICATION NUMBER: US SN 07/956,700  
FILING DATE: 02-OCT-1992  
CLASSIFICATION: 800  
APPLICATION NUMBER: PCT/US93/09340  
FILING DATE: 30-SEP-1993  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Kitchell, Barbara S.  
REGISTRATION NUMBER: 33,928  
REFERENCE/DOCKET NUMBER: ARCD:152/KIT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 453 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-468-793-8

Query Match 17.1%; Score 1005.5; DB 4; Length 453;  
Best Local Similarity 47.1%; Pred. No. 2e-71;  
Matches 209; Conservative 72; Mismatches 158; Indels 5; Gaps 3;  
QY 29 FKKILVANRGEIAVRAFAALETGAATVAIYPREDRGSFHRSEAFSEAVRIGTEGSPVKAY 88  
DB 3 FNKILIANRGEIALRILRTCEELGIGTIAVHSTVDNRNALHVQLADEAVCIG-EAASSKSY 61  
QY 89 LDIDEIIGAARKVKADAIYPGYGFLSENAOLARECAENGITFTIGPTPEVLDLTGDKSRV 148  
DB 62 LNIPIIAAALTRNASAIHPGYGFLAENARFAEICADHHLTFIGSPDSIRAMGDKSTAK 121  
QY 149 TAACKAGLPVLAESTP-SKNIDDIVKSAEGQYPIFVKAVAGGGGMRVSSPDELRLK 207  
DB 122 ETMQRVGPTIPGSDGLLTDVDSAAKVAEIGYPMIKATAGGGGMRVLPREPADLEK 181  
QY 208 ATEASRAEAFAFGDSVYVERAVINPOHIEVQILGDRTEGVVHLVYERDCSLORRHOKVVE 267  
DB 182 FLAAQGEAEAFGNPGLYLEKIDRPRHVEGQILADAYGNVHVGDERDCSIORRHOKLLE 241  
QY 268 IAPAHLDPELDRICADAVKFCRSIGYQAGTVFELVDEKGNHVFIEPNRIQVEHTVT 327  
DB 242 EAPSPALSADLRQMGDAAVKVAQAIYIGAGTVFELVDATGNTFYFEMNTRIQVEHPVT 301  
QY 328 EEVTEVDLVKAQMLAAGATLKEGLTQDKIKTHGAALOCRTITTEDPNNRFRPDTGTITA 387  
DB 302 EMITGLDLIAEQIRIAQGEALR---FRQADIQLRGAIECRINAEDPEYFNPNRPGRTG 358  
QY 388 YRSPGGAGVRLDGAALGGGEITAHFDSMLVKMTCRGSDFETAVARAQALAEFTVSGVAT 447  
DB 359 YLPPGGPGVRVDSHVYTDYEIPPYDLSLIGKLIWVGATREAIARMQALRECAITGLPT 418  
QY 448 NIGFLRALLREEDFTSKRIATGFI 471

RESULT 13  
US-07-956-700B-6  
Sequence 6, Application US/07956700B  
Patent No. 5539092  
GENERAL INFORMATION:  
APPLICANT: Robert Haselkorn and Piotr Gornicki

DB 419 TLSFHOLMLQMPFELRGELYTNFV 442  
RESULT 12  
US-09-433-043B-121  
Sequence 121, Application US/09433043B  
Patent No. 6399342  
GENERAL INFORMATION:  
APPLICANT: HASSELKORN, ROBERT  
APPLICANT: GORNICKI, PIOTR  
TITLE OF INVENTION: CYANOBACTERIAL AND PLANT ACETYL-CoA CARBOXYLASE  
FILE REFERENCE: ARCD:338US  
CURRENT APPLICATION NUMBER: US/09/433,043B  
CURRENT FILING DATE: 1999-10-25  
PRIOR APPLICATION NUMBER: 08/475,879  
PRIOR FILING DATE: 1995-06-07  
PRIOR APPLICATION NUMBER: 07/956,700  
PRIOR FILING DATE: 1992-10-02  
NUMBER OF SEQ ID NOS: 128  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO: 121  
LENGTH: 453  
TYPE: PPT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
OTHER INFORMATION: Peptide  
US-09-433-043B-121

Query Match 17.1%; Score 1003.5; DB 4; Length 453;  
Best Local Similarity 47.1%; Pred. No. 2.9e-71;  
Matches 209; Conservative 72; Mismatches 158; Indels 5; Gaps 3;  
QY 29 FKKILVANRGEIAVRAFAALETGAATVAIYPREDRGSFHRSEAFSEAVRIGTEGSPVKAY 88  
DB 3 FNKILIANRGEIALRILRTCEELGIGTIAVHSTVDNRNALHVQLADEAVCIG-EAASSKSY 61  
QY 89 LDIDEIIGAARKVKADAIYPGYGFLSENAOLARECAENGITFTIGPTPEVLDLTGDKSRV 148  
DB 62 LNIPIIAAALTRNASAIHPGYGFLAENARFAEICADHHLTFIGSPDSIRAMGDKSTAK 121  
QY 149 TAACKAGLPVLAESTP-SKNIDDIVKSAEGQYPIFVKAVAGGGGMRVSSPDELRLK 207  
DB 122 ETMQRVGPTIPGSDGLLTDVDSAAKVAEIGYPMIKATAGGGGMRVLPREPADLEK 181  
QY 208 ATEASRAEAFAFGDSVYVERAVINPOHIEVQILGDRTEGVVHLVYERDCSLORRHOKVVE 267  
DB 182 FLAAQGEAEAFGNPGLYLEKIDRPRHVEGQILADAYGNVHVGDERDCSIORRHOKLLE 241  
QY 268 IAPAHLDPELDRICADAVKFCRSIGYQAGTVFELVDEKGNHVFIEPNRIQVEHTVT 327  
DB 242 EAPSPALSADLRQMGDAAVKVAQAIYIGAGTVFELVDATGNTFYFEMNTRIQVEHPVT 301  
QY 328 EEVTEVDLVKAQMLAAGATLKEGLTQDKIKTHGAALOCRTITTEDPNNRFRPDTGTITA 387  
DB 302 EMITGLDLIAEQIRIAQGEALR---FRQADIQLRGAIECRINAEDPEYFNPNRPGRTG 358  
QY 388 YRSPGGAGVRLDGAALGGGEITAHFDSMLVKMTCRGSDFETAVARAQALAEFTVSGVAT 447  
DB 359 YLPPGGPGVRVDSHVYTDYEIPPYDLSLIGKLIWVGATREAIARMQALRECAITGLPT 418  
QY 448 NIGFLRALLREEDFTSKRIATGFI 471  
DB 419 TLSFHOLMLQMPFELRGELYTNFV 442

RESULT 13  
US-07-956-700B-6  
Sequence 6, Application US/07956700B  
Patent No. 5539092  
GENERAL INFORMATION:  
APPLICANT: Robert Haselkorn and Piotr Gornicki

;; TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA  
;; TITLE OF INVENTION: Carboxylase  
;; NUMBER OF SEQUENCES: 116  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Arnold, White & Durkee  
;; STREET: 321 No. 553092th Clark Street  
;; CITY: Chicago  
;; STATE: Illinois  
;; COUNTRY: USA  
;; ZIP: 60610  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy Disk  
;; COMPUTER: IBM PC Compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: ASCII-DOS  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/07/956,700B  
;; FILING DATE: 19921002  
;; CLASSIFICATION: 800  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Thomas E. No. 553092thrup  
;; REGISTRATION NUMBER: 33,268  
;; REFERENCE/DOCKET NUMBER: ARCD:058  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 1-312-744-0090  
;; TELEFAX: 1-312-755-4489  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 453 amino acids  
;; TYPE: Amino acid  
;; STRANDEDNESS: Single  
;; TOPOLOGY: Linear  
;; MOLECULE TYPE: Peptide  
;; US-07-956-700B-6

Query Match 17.0%; Score 996.5; DB 1; Length 453;  
Best Local Similarity 46.8%; Pred. No. 1e-70;  
Matches 208; Conservative 72; Mismatches 159; Indels 5; Gaps 3;

QY 29 FKILVANRGEIAVRAFAALETGAATVAIYPREDRSGFHRSEAFSEAVRIGTEGSPVKAY 88  
Db 3 FNKILIANRGEIALRILRCEELGIGTIAVHSTVDNRNALHVLADEAVCIG-EAASSKSY 61  
QY 89 LDIDEIIGAIAKKVADAIYPGYGFLSENAQLARECAENGITFGTPEVLDLTGDKSRV 148  
Db 62 LNPINIIAALTRNASHIHGYGFLAENARFAEICADHHLTFITGSPDSIRAMGDKSTAK 121  
QY 149 TAAKKAGLPVLAESTP-SKNIDDIVKSAEGQTPYIFVKAVAGGGGRMRVSSPDELRLK 207  
Db 122 ETMQRVGVPTIPGSDGLLTVDVSAAKVAEEIGYPMVIMIKATAGGGGRMRLVREPADLEK 181  
QY 208 ATEASREAEAAFGDGSVYVERAVINPOHIEVQILGDRTEGVVHLYERDCSLORRHOKVVE 267  
Db 182 FLAQQEAEAAFGDGSVYVERAVINPOHIEVQILGDRTEGVVHLYERDCSLORRHOKVVE 241  
QY 268 IAPAHQLDPELRICADAVKFCRSICYGAGTVEFLVDEKGNHVFTEMPNRIQVEHTVT 327  
Db 242 EAPSPALSADLRQKMGDAAVKVAQAIGYIGAGTVEFLVDTGATGNFYFNMNTRIOVEHPT 301  
QY 328 EEVTEVDLVKAQMLAAGATLTKELGILTQDKIKTHGAALQCRITTEDPNNNGFRPDTGITTA 387  
Db 302 EMITGLDLIAEQIRIAQGEALR---FRQADIQLRGHAEICRINAEDPEYFNPGRITG 358  
QY 388 YRSPGAGVPLDGAALGGEITAHFDSMLVKMTCRGSDPETAVARAORALAEFTVSGVAT 447  
Db 359 YLPPGGVGVKVDVSHVYTDYIPYDLSLIGKLLVWGATREEARQORALREGAITGLPT 418  
QY 448 NIGFALLREEDFTSKRIATGFI 471  
Db 419 TLSFHQLMLQMPPEFLRGELYTNEV 442

RESULT 14

US-08-476-537-6  
;; Sequence 6, Application US/08476537  
;; Patent No. 5756290  
;; GENERAL INFORMATION:  
;; APPLICANT: Robert Haselkorn and Piotr Gornicki  
;; TITLE OF INVENTION: Cyanobacterial and Plant Acetyl-CoA  
;; TITLE OF INVENTION: Carboxylase  
;; NUMBER OF SEQUENCES: 116  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Arnold, White & Durkee  
;; STREET: 321 No. 5756290th Clark Street  
;; CITY: Chicago  
;; STATE: Illinois  
;; COUNTRY: USA  
;; ZIP: 60610  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy Disk  
;; COMPUTER: IBM PC Compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: ASCII-DOS  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/476,537  
;; FILING DATE:  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: 07/956,700  
;; FILING DATE: 10/21/92  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Thomas E. No. 5756290thrup  
;; REGISTRATION NUMBER: 33,268  
;; REFERENCE/DOCKET NUMBER: ARCD:058  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 1-312-744-0090  
;; TELEFAX: 1-312-755-4489  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 453 amino acids  
;; TYPE: Amino acid  
;; STRANDEDNESS: Single  
;; TOPOLOGY: Linear  
;; MOLECULE TYPE: Peptide  
;; US-08-476-537-6

Query Match 17.0%; Score 996.5; DB 1; Length 453;  
Best Local Similarity 46.8%; Pred. No. 1e-70;  
Matches 208; Conservative 72; Mismatches 159; Indels 5; Gaps 3;

QY 29 FKILVANRGEIAVRAFAALETGAATVAIYPREDRSGFHRSEAFSEAVRIGTEGSPVKAY 88  
Db 3 FNKILIANRGEIALRILRCEELGIGTIAVHSTVDNRNALHVLADEAVCIG-EAASSKSY 61  
QY 89 LDIDEIIGAIAKKVADAIYPGYGFLSENAQLARECAENGITFGTPEVLDLTGDKSRV 148  
Db 62 LNPINIIAALTRNASHIHGYGFLAENARFAEICADHHLTFITGSPDSIRAMGDKSTAK 121  
QY 149 TAAKKAGLPVLAESTP-SKNIDDIVKSAEGQTPYIFVKAVAGGGGRMRVSSPDELRLK 207  
Db 122 ETMQRVGVPTIPGSDGLLTVDVSAAKVAEEIGYPMVIMIKATAGGGGRMRLVREPADLEK 181  
QY 208 ATEASREAEAAFGDGSVYVERAVINPOHIEVQILGDRTEGVVHLYERDCSLORRHOKVVE 267  
Db 182 FLAQQEAEAAFGDGSVYVERAVINPOHIEVQILGDRTEGVVHLYERDCSLORRHOKVVE 241  
QY 268 IAPAHQLDPELRICADAVKFCRSICYGAGTVEFLVDEKGNHVFTEMPNRIQVEHTVT 327  
Db 242 EAPSPALSADLRQKMGDAAVKVAQAIGYIGAGTVEFLVDTGATGNFYFNMNTRIOVEHPT 301  
QY 328 EEVTEVDLVKAQMLAAGATLTKELGILTQDKIKTHGAALQCRITTEDPNNNGFRPDTGITTA 387  
Db 302 EMITGLDLIAEQIRIAQGEALR---FRQADIQLRGHAEICRINAEDPEYFNPGRITG 358  
QY 388 YRSPGAGVPLDGAALGGEITAHFDSMLVKMTCRGSDPETAVARAORALAEFTVSGVAT 447



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 19:02:46 ; Search time 24 Seconds  
(without alignments)  
2830.970 Million cell updates/sec

Title: US-09-974-973-2

Perfect score: 5865

Sequence: 1 MTAITLGGLLKGIIITVST.....RVVVPAAFKVEGGDLIVVVS 1157

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5865	100.0	1157	9	US-09-974-973-2
2	5865	100.0	1157	9	US-09-974-973-4
3	5759	98.2	1140	9	US-09-974-973-19
4	5759	98.2	1140	9	US-09-738-626-4265
5	5759	98.2	1140	9	US-10-045-072-2
6	2489	42.4	1147	10	US-09-815-242-5468
7	2460.5	42.0	1142	10	US-09-815-242-10806
8	2300	39.2	1073	10	US-09-815-242-12361
9	1032.5	17.6	447	10	US-09-767-479-6
10	1005.5	17.1	453	10	US-09-767-479-8
11	979.5	16.7	471	10	US-09-815-242-5215
12	956	16.3	449	10	US-09-815-242-13885
13	950	16.2	448	10	US-09-815-242-11160
14	940	16.0	449	10	US-09-815-242-10330
15	936.5	16.0	455	10	US-09-815-242-13617
16	935.5	16.0	455	10	US-09-815-242-13364
17	921	15.7	449	10	US-09-815-242-12063
18	919	15.7	458	10	US-09-815-242-11321
19	909	15.5	443	10	US-09-815-242-4963

20	909	15.5	456	10	US-09-815-242-10924	Sequence 10324, A
21	902	15.4	455	10	US-09-815-242-11558	Sequence 11558, A
22	880.5	15.0	591	9	US-09-738-626-6940	Sequence 6940, Ap
23	856.5	14.6	446	10	US-09-815-242-5418	Sequence 5418, Ap
24	856.5	14.6	453	10	US-09-815-242-12562	Sequence 12562, A
25	837	14.3	448	10	US-09-815-242-5806	Sequence 5806, Ap
26	834	14.2	451	10	US-09-815-242-13127	Sequence 13127, A
27	633	10.8	358	10	US-09-815-242-12939	Sequence 12939, A
28	583.5	9.9	2257	10	US-09-767-479-10	Sequence 10, Appl
29	437.5	7.5	483	9	US-10-083-357-1328	Sequence 1328, Ap
30	377.5	6.4	158	10	US-09-815-242-5031	Sequence 5031, Ap
31	300.5	5.1	163	9	US-09-895-913A-324	Sequence 324, App
32	251	4.3	124	10	US-09-205-658-240	Sequence 240, App
33	244.5	4.2	1687	9	US-10-094-679-3	Sequence 3, Appli
34	225	3.8	124	10	US-09-205-658-238	Sequence 238, App
35	213.5	3.6	1113	9	US-09-738-626-5279	Sequence 5279, Ap
36	208.5	3.6	98	10	US-09-815-242-12938	Sequence 12938, A
37	208.5	3.6	1113	10	US-09-836-470B-3	Sequence 3, Appli
38	185	3.2	4999	9	US-09-976-059-15	Sequence 15, Appl
39	170	2.9	1289	9	US-09-712-363-259	Sequence 259, App
40	165.5	2.8	2969	9	US-09-738-626-4434	Sequence 4434, Ap
41	163.5	2.8	5215	9	US-09-860-846-2	Sequence 2, Appli
42	163.5	2.8	5215	10	US-09-861-289-2	Sequence 2, Appli
43	162.5	2.8	122	10	US-09-987-485-1	Sequence 1, Appli
44	160	2.7	4999	9	US-09-976-059-14	Sequence 14, Appl
45	158.5	2.7	70	10	US-09-987-485-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-974-973-2  
; Sequence 2, Application US/09974973  
; Patent No. US20020177202A1  
; GENERAL INFORMATION:  
; APPLICANT: Hanke, Paul D.  
; TITLE OF INVENTION: Feedback-Resistant Pyruvate Carboxylase Gene from Corynebacter  
; FILE REFERENCE: 1533.1230001/MAC/RGM  
; CURRENT APPLICATION NUMBER: US/09/974,973  
; CURRENT FILING DATE: 2001-10-21  
; PRIOR APPLICATION NUMBER: US 60/239,913  
; PRIOR FILING DATE: 2000-10-13  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 1157  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-974-973-2

Query Match	100.0%	Score 5865;	DB 9;	Length 1157;
Best Local Similarity	100.0%	Pred. No. 0;	Mismatches 0;	Indels 0;
Matches 1157;	Conservative	0;		Gaps 0;
Qy	1	MTAITLGGLLKGIIITVSTHTSSTLPAPFKKILVANRGEIAVRAALETGAATVAIYP	60	
Db	1	MTAITLGGLLKGIIITVSTHTSSTLPAPFKKILVANRGEIAVRAALETGAATVAIYP	60	
Qy	61	REDRGSFHRSFASEAVRICTEGSPVKAYLDIDEIIGAARKVKADATYPGYFLSENAQLA	120	
Db	61	REDRGSFHRSFASEAVRICTEGSPVKAYLDIDEIIGAARKVKADATYPGYFLSENAQLA	120	
Qy	121	RECAENGITFIPTPEVLDLTGDKSRVTAANKAGLPLVLAESTPSKNIDDIIVKSAEQTY	180	
Db	121	RECAENGITFIPTPEVLDLTGDKSRVTAANKAGLPLVLAESTPSKNIDDIIVKSAEQTY	180	
Qy	181	PIFVKAVAGGGGRMRFVSPDELRLKLA TEASREAAAFGDGSVYVERAVINPQHLEVOI	240	
Db	181	PIFVKAVAGGGGRMRFVSPDELRLKLA TEASREAAAFGDGSVYVERAVINPQHLEVOI	240	
Qy	241	LGDRTEGVVHLRYDCSLQRRHQKVVEIAPAQHLDELRLDRICADAVKFCRSIGYGAGT	300	
Db	241	LGDRTEGVVHLRYDCSLQRRHQKVVEIAPAQHLDELRLDRICADAVKFCRSIGYGAGT	300	

Db 241 LGDRTGEVHVLYERDCSLQRHQKVVETAPQAHLDPDLRDRICADAVKFCRSIGYQAGT 300  
Qy 301 VEFVLDEKGNHVFTEMNPRIOVEHTVEEVDLVKAQMLAAGATLKEIGLTDQIKT 360  
Db 301 VEFVLDEKGNHVFTEMNPRIOVEHTVEEVDLVKAQMLAAGATLKEIGLTDQIKT 360  
Qy 361 HGAALOCRIITTEDPNNGFRPDGTITAYRSPGGAGVRLDGAALQGGEITAHFDSMLVKMT 420  
Db 361 HGAALOCRIITTEDPNNGFRPDGTITAYRSPGGAGVRLDGAALQGGEITAHFDSMLVKMT 420  
Qy 421 CRGSDFETAVARAQALAEFTVSGVATNIGFLRALLREEDFTSKRIATGFTGDHPLLQA 480  
Db 421 CRGSDFETAVARAQALAEFTVSGVATNIGFLRALLREEDFTSKRIATGFTGDHPLLQA 480  
Qy 481 PPADDEQGRILDLADVTYNKPHGVKDVAAAPDKLPNIKDLPLPGSRDLKQLGPAA 540  
Db 481 PPADDEQGRILDLADVTYNKPHGVKDVAAAPDKLPNIKDLPLPGSRDLKQLGPAA 540  
Qy 541 FARDLREQDALAVTDTTFRDAHQSLATRVRSFALKPAAEAVAKLTPELLSVEAWGATY 600  
Db 541 FARDLREQDALAVTDTTFRDAHQSLATRVRSFALKPAAEAVAKLTPELLSVEAWGATY 600  
Qy 601 DVAMRFLFEDPMDRLDELREAMPNWNOMLLRGNTVGYTYPDSVCRAFVKEAASGYD 660  
Db 601 DVAMRFLFEDPMDRLDELREAMPNWNOMLLRGNTVGYTYPDSVCRAFVKEAASGYD 660  
Qy 661 IFRIFDALNDVSQMRPAIDAVLETNTAVAEVAMAYSGDLSDPNEKLYTLDYILKMAEIV 720  
Db 661 IFRIFDALNDVSQMRPAIDAVLETNTAVAEVAMAYSGDLSDPNEKLYTLDYILKMAEIV 720  
Qy 721 KSGAHILAIDMAGLLRPAAVTKLVTLRREFDLPVHVHTDTAGGOLATYFAAAQAGAD 780  
Db 721 KSGAHILAIDMAGLLRPAAVTKLVTLRREFDLPVHVHTDTAGGOLATYFAAAQAGAD 780  
Qy 781 AVDGASAPLSGTTTSPSLSAIVAAFAHTRDRTGLSLEAVSDLEPYWEAVRGLYLPESGT 840  
Db 781 AVDGASAPLSGTTTSPSLSAIVAAFAHTRDRTGLSLEAVSDLEPYWEAVRGLYLPESGT 840  
Qy 841 PGPTGRVYRHEIPGQSLNLRQAATLGLADREFELIEDNVAANMELGRPTKVTTPSSKV 900  
Db 841 PGPTGRVYRHEIPGQSLNLRQAATLGLADREFELIEDNVAANMELGRPTKVTTPSSKV 900  
Qy 901 GDALHLVGAGVDPADFAADPKYDIIPDSVIAFLRGELGNPPGWPPEPLTRALEGRSEG 960  
Db 901 GDALHLVGAGVDPADFAADPKYDIIPDSVIAFLRGELGNPPGWPPEPLTRALEGRSEG 960  
Qy 961 KAPLTEVPEEQAHLDADDKSKERNNSLRLLPKPTPEEFLEHRRRFGNTSALDDREFFYG 1020  
Db 961 KAPLTEVPEEQAHLDADDKSKERNNSLRLLPKPTPEEFLEHRRRFGNTSALDDREFFYG 1020  
Qy 1021 LVEGRETLIRLPDVRTPLLRDLAISPPDDKGRNVVANYNGQIRPMRVDRSVESVTAT 1080  
Db 1021 LVEGRETLIRLPDVRTPLLRDLAISPPDDKGRNVVANYNGQIRPMRVDRSVESVTAT 1080  
Qy 1081 AEKADSSNKGHVAAAPFAGVVTVTVAEGDEVKAGDAVAIIEAMKMEATITASVDGKIERYV 1140  
Db 1081 AEKADSSNKGHVAAAPFAGVVTVTVAEGDEVKAGDAVAIIEAMKMEATITASVDGKIERYV 1140  
Qy 1141 VPAATKVEGGDLIVVVS 1157  
Db 1141 VPAATKVEGGDLIVVVS 1157

## RESULT 2

US-09-974-973-4  
; Sequence 4, Application us/09974973  
; Patent No. US20020171202A1  
; GENERAL INFORMATION:  
; APPLICANT: Hanke, Paul D.  
; TITLE OF INVENTION: Feedback-Resistant Pyruvate Carboxylase Gene from *Corynebacterium*  
; FILE REFERENCE: 1533-1230001/MAC/RGM  
; CURRENT APPLICATION NUMBER: US/09/974, 973  
; CURRENT FILING DATE: 2001-10-21

; PRIOR APPLICATION NUMBER: US 60/239,913  
; PRIOR FILING DATE: 2000-10-13  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 4  
; LENGTH: 1157  
; TYPE: PRT  
; ORGANISM: *Corynebacterium glutamicum*  
US-09-974-973-4

## Query Match

Best Local Similarity 100.0%; Score 5865; DB 9; Length 1157;  
Matches 1157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTAITLGGLLKGIITLVSTHTSSSTLPAPKKILVANRGEIAVRAALETGAATVAIYP 60

Db 1 MTAITLGGLLKGIITLVSTHTSSSTLPAPKKILVANRGEIAVRAALETGAATVAIYP 60

Qy 61 REDRGSFHRSEAVRIGTEGSPVKAYLIDIEIIGAARKVKADAIYPGYFLSENAQLA 120

Db 61 REDRGSFHRSEAVRIGTEGSPVKAYLIDIEIIGAARKVKADAIYPGYFLSENAQLA 120

Qy 121 RECAENGITF IGTPPEVLDTGDKSRVTAACKAGLPVLAESTPSKNIDDIIVKSAEGQY 180

Db 121 RECAENGITF IGTPPEVLDTGDKSRVTAACKAGLPVLAESTPSKNIDDIIVKSAEGQY 180

Qy 181 PIFVKAVAGGGGRMRFVSSPDRLKLAETASREAAAFDGSVYVERAVINPOHIEVOI 240

Db 181 PIFVKAVAGGGGRMRFVSSPDRLKLAETASREAAAFDGSVYVERAVINPOHIEVOI 240

Qy 241 LGDRTGEVHVLYERDCSLQRHQKVVETAPQAHLDPDLRDRICADAVKFCRSIGYQAGT 300

Db 241 LGDRTGEVHVLYERDCSLQRHQKVVETAPQAHLDPDLRDRICADAVKFCRSIGYQAGT 300

Qy 301 VEFVLDEKGNHVFTEMNPRIOVEHTVEEVDLVKAQMLAAGATLKEIGLTDQIKT 360

Db 301 VEFVLDEKGNHVFTEMNPRIOVEHTVEEVDLVKAQMLAAGATLKEIGLTDQIKT 360

Qy 361 HGAALOCRIITTEDPNNGFRPDGTITAYRSPGGAGVRLDGAALQGGEITAHFDSMLVKMT 420

Db 361 HGAALOCRIITTEDPNNGFRPDGTITAYRSPGGAGVRLDGAALQGGEITAHFDSMLVKMT 420

Qy 421 CRGSDFETAVARAQALAEFTVSGVATNIGFLRALLREEDFTSKRIATGFTGDHPLLQA 480

Db 421 CRGSDFETAVARAQALAEFTVSGVATNIGFLRALLREEDFTSKRIATGFTGDHPLLQA 480

Qy 481 PPADDEQGRILDLADVTYNKPHGVKDVAAAPDKLPNIKDLPLPGSRDLKQLGPAA 540

Db 481 PPADDEQGRILDLADVTYNKPHGVKDVAAAPDKLPNIKDLPLPGSRDLKQLGPAA 540

Qy 541 FARDLREQDALAVTDTTFRDAHQSLATRVRSFALKPAAEAVAKLTPELLSVEAWGATY 600

Db 541 FARDLREQDALAVTDTTFRDAHQSLATRVRSFALKPAAEAVAKLTPELLSVEAWGATY 600

Qy 601 DVAMRFLFEDPMDRLDELREAMPNWNOMLLRGNTVGYTYPDSVCRAFVKEAASGYD 660

Db 601 DVAMRFLFEDPMDRLDELREAMPNWNOMLLRGNTVGYTYPDSVCRAFVKEAASGYD 660

Qy 661 IFRIFDALNDVSQMRPAIDAVLETNTAVAEVAMAYSGDLSDPNEKLYTLDYILKMAEIV 720

Db 661 IFRIFDALNDVSQMRPAIDAVLETNTAVAEVAMAYSGDLSDPNEKLYTLDYILKMAEIV 720

Qy 721 KSGAHILAIDMAGLLRPAAVTKLVTLRREFDLPVHVHTDTAGGOLATYFAAAQAGAD 780

Db 721 KSGAHILAIDMAGLLRPAAVTKLVTLRREFDLPVHVHTDTAGGOLATYFAAAQAGAD 780

Qy 781 AVDGASAPLSGTTTSPSLSAIVAAFAHTRDRTGLSLEAVSDLEPYWEAVRGLYLPESGT 840

Db 781 AVDGASAPLSGTTTSPSLSAIVAAFAHTRDRTGLSLEAVSDLEPYWEAVRGLYLPESGT 840

Qy 841 PGPTGRVYRHEIPGQSLNLRQAATLGLADREFELIEDNVAANMELGRPTKVTTPSSKV 900

Db 841 PGPTGRVYRHEIPGQSLNLRQAATLGLADREFELIEDNVAANMELGRPTKVTTPSSKV 900



QY 901 GDALHLVAGVDPADPAADPKYDIPDSVIAFLRGLGNPPGPGWPEPLRTRALEGRSEG 960  
Db 901 GDALHLVAGVDPADPAADPKYDIPDSVIAFLRGLGNPPGPGWPEPLRTRALEGRSEG 960  
QY 961 KAPLTPVEPEEQAHLDADDDSKERNLSNRLFFPKPTPEEFLEHRRRFGNTSALDDREFFYG 1020  
Db 961 KAPLTPVEPEEQAHLDADDDSKERNLSNRLFFPKPTPEEFLEHRRRFGNTSALDDREFFYG 1020  
QY 1021 LVSEGRETLIRLPDVRTPELLVRLDAISEPDDKGMNVVYVNGQIRPMRVDRDRSVESVTAT 1080  
Db 1021 LVSEGRETLIRLPDVRTPELLVRLDAISEPDDKGMNVVYVNGQIRPMRVDRDRSVESVTAT 1080  
QY 1081 AERADSSNKGHVAAPFAGVYTVVVAEGDEVKAGDAVAIIIEAMKMEATITASVDGKTERVV 1140  
Db 1081 AERADSSNKGHVAAPFAGVYTVVVAEGDEVKAGDAVAIIIEAMKMEATITASVDGKTERVV 1140  
QY 1141 VPAATKVEGGDLIVVVS 1157  
Db 1141 VPAATKVEGGDLIVVVS 1157  
RESULT 3  
US-09-974-973-19  
; Sequence 19, Application US/09974973  
; Patent No. US20020177202A1  
; GENERAL INFORMATION:  
; APPLICANT: Hanke, Paul D.  
; TITLE OF INVENTION: Feedback-Resistant Pyruvate Carboxylase Gene from *Corynebacterium*  
; FILE REFERENCE: 1533.1230001/MAC/RGM  
; CURRENT APPLICATION NUMBER: US/09/974,973  
; CURRENT FILING DATE: 2001-10-21  
; PRIOR APPLICATION NUMBER: US 60/239,913  
; PRIOR FILING DATE: 2000-10-13  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 19  
; LENGTH: 1140  
; TYPE: PRT  
; ORGANISM: *Corynebacterium glutamicum*  
US-09-974-973-19  
Query Match 98.2%; Score 5759; DB 9; Length 1140;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1133; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
QY 18 VSPHTSTLPAPFKKILVANGETAVRAALETGAATVAIYPREDGSGFHRSEAVR 77  
Db 1 MSTHTSTLPAPFKKILVANGETAVRAALETGAATVAIYPREDGSGFHRSEAVR 60  
QY 78 IGTEGSPVKAYLDIDEITIGAARKVKADAIYPGYGFLSENAQLARECAENGITFGTPEV 137  
Db 61 IGTEGSPVKAYLDIDEITIGAARKVKADAIYPGYGFLSENAQLARECAENGITFGTPEV 120  
QY 138 LDTGDKSRVATAAKKAGLPVLAESTPSKNIDDIVSAEGQTYPIFVKAVAGGGRGMRF 197  
Db 121 LDTGDKSRVATAAKKAGLPVLAESTPSKNIDDIVSAEGQTYPIFVKAVAGGGRGMRF 180  
QY 198 VSSPDELRLKLTASREAEAAFGDGSVYVERAVINPQHIEVOILGDBTGEVHLYERDCS 257  
Db 181 VASPDRLKLTASREAEAAFGDGSVYVERAVINPQHIEVOILGDBTGEVHLYERDCS 240  
QY 258 LQRRHKQVVEIAPQAHLDPRLDRICADAVKFCRSIGYQAGTVFELVDEKGNHVFTEMN 317  
Db 241 LQRRHKQVVEIAPQAHLDPRLDRICADAVKFCRSIGYQAGTVFELVDEKGNHVFTEMN 300  
QY 318 PRIOQVHTVTEVTEVDLVKAQMLAAGATLTKELGLTQDKIKTHGAALQCRITTEDPNNG 377  
Db 301 PRIOQVHTVTEVTEVDLVKAQMLAAGATLTKELGLTQDKIKTHGAALQCRITTEDPNNG 360  
QY 378 FRPDTGTITAYRSPGGAGVRLDGAALGGSEITAHFDSMLVKMTCRGSDFETAVARAQAL 437  
Db 361 FRPDTGTITAYRSPGGAGVRLDGAALGGSEITAHFDSMLVKMTCRGSDFETAVARAQAL 420

QY 438 AETVSCVATNIGFLRALLREEDFTSKRIATGFTGDHPHLLQAPPADDEOGRILDYLDV 497  
Db 421 AETVSCVATNIGFLRALLREEDFTSKRIATGFTADHPHLLQAPPADDEOGRILDYLDV 480  
QY 498 TVNKPCHVVRPKDVAAPIDKLPNIKDLPLPRGSRDLKQLGPAAPARDLREQDALAVTDTT 557  
Db 481 TVNKPCHVVRPKDVAAPIDKLPNIKDLPLPRGSRDLKQLGPAAPARDLREQDALAVTDTT 540  
QY 558 FRAHQSLLATRVRSFALKPAAEAVAKLTPELLSVEAWGATYDVAMRFLFEDPWRDLE 617  
Db 541 FRAHQSLLATRVRSFALKPAAEAVAKLTPELLSVEAWGATYDVAMRFLFEDPWRDLE 600  
QY 618 LRAMPNVNTQMLLGRNTVGYTYPYDSCVCRAFVKEAASSGVDFIRFDALNDVQMRPA 677  
Db 601 LRAMPNVNTQMLLGRNTVGYTYPYDSCVCRAFVKEAASSGVDFIRFDALNDVQMRPA 660  
QY 678 IDAVLETNTAAVEAMAYSGDLSDPNEKLYTLDYLLKMAEEIVKSGAHILAIKDMAGLLR 737  
Db 661 IDAVLETNTAAVEAMAYSGDLSDPNEKLYTLDYLLKMAEEIVKSGAHILAIKDMAGLLR 720  
QY 738 PAAVTKLVTALRRREFDLPVHVHTHDTAGGQLATYFAAAQAGADAVDGSAPLSGTTSQPS 797  
Db 721 PAAVTKLVTALRRREFDLPVHVHTHDTAGGQLATYFAAAQAGADAVDGSAPLSGTTSQPS 780  
QY 798 LSAIVAAFAHTRDRTGLSLLEAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGQL 857  
Db 781 LSAIVAAFAHTRDRTGLSLLEAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGQL 840  
QY 858 SNLRAQATALGLADREFELIEDNYAAVNEMLGRPTKVTSSKVVGDALHLVAGVDPADF 917  
Db 841 SNLRAQATALGLADREFELIEDNYAAVNEMLGRPTKVTSSKVVGDALHLVAGVDPADF 900  
QY 918 AADPKYDIPDSVIAFLRGLGNPPGPGWPEPLRTRALEGRSEKAPLTPVEPEEQAHLD 977  
Db 901 AADPKYDIPDSVIAFLRGLGNPPGPGWPEPLRTRALEGRSEKAPLTPVEPEEQAHLD 960  
QY 978 DSKERNLSNRLFFPKPTPEEFLEHRRRFGNTSALDDREFFYGLVEGRETLIRLPDVRTP 1037  
Db 961 DSKERNLSNRLFFPKPTPEEFLEHRRRFGNTSALDDREFFYGLVEGRETLIRLPDVRTP 1020  
QY 1038 LLVRLDAISEPDDKGMNVVYVNGQIRPMRVDRDRSVESVTATAEKADSSNKGHVAAAPFA 1097  
Db 1021 LLVRLDAISEPDDKGMNVVYVNGQIRPMRVDRDRSVESVTATAEKADSSNKGHVAAAPFA 1080  
QY 1098 GVVTVTVAEDEVKAGDAVAIIIEAMKMEATITASVDGKIERVVVPAATKVEGGDLIVVVS 1157  
Db 1081 GVVTVTVAEDEVKAGDAVAIIIEAMKMEATITASVDGKIDRVVVVPAATKVEGGDLIVVVS 1140  
RESULT 4  
US-09-738-626-4265  
; Sequence 4265, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07

; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 4265  
; LENGTH: 1140  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4265

Query Match 98.2%; Score 5759; DB 9; Length 1140;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1133; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 18 VSTHTSSTLPAPFKKILVANRGEIARAFRAALETGAATVAIYPREDRGSFHRSEAVR 77  
Db 1 MSTHTSSTLPAPFKKILVANRGEIARAFRAALETGAATVAIYPREDRGSFHRSEAVR 60

QY 78 IGTEGSPVKAYLIDIDEIIGAARKKADAIYPGYFLSENAQLARECAENGITFIGPTPEV 137  
Db 61 IGTEGSPVKAYLIDIDEIIGAARKKADAIYPGYFLSENAQLARECAENGITFIGPTPEV 120

QY 138 LDLTGDKSRANTAAKAGLPVLAESTPSKNIDDIIVKSAEGQTYPIFVKAVAGGGGRMRF 197  
Db 121 LDLTGDKSRANTAAKAGLPVLAESTPSKNIDDIIVKSAEGQTYPIFVKAVAGGGGRMRF 180

QY 198 YSSDELKRLATEASREAAFGDGVYVERAVINPQHIIEVQILGDRGTGEVHHLYERDCS 257  
Db 181 VASPDRLKRLATEASREAAFGDGVYVERAVINPQHIIEVQILGDRGTGEVHHLYERDCS 240

QY 258 LQRRHQKVEIAPAOHLDPRLDRICADAVKFCRSIGYQAGTVEFLVDEKGNHVFTEMN 317  
Db 241 LQRRHQKVEIAPAOHLDPRLDRICADAVKFCRSIGYQAGTVEFLVDEKGNHVFTEMN 300

QY 318 PRIQVEHTVTEVTEDVLVKAQMRLAAGATLKELGLTQDKIKTHGAALOCRTITEDPNNG 377  
Db 301 PRIQVEHTVTEVTEDVLVKAQMRLAAGATLKELGLTQDKIKTHGAALOCRTITEDPNNG 360

QY 378 FRPDGTITAYRSPGAGVRLDGAOLGGEITAHFDSMLVKMTCRGSDFETAVARAQAL 437  
Db 361 FRPDGTITAYRSPGAGVRLDGAOLGGEITAHFDSMLVKMTCRGSDFETAVARAQAL 420

QY 438 AEFTVSGVATNIGFRLALREEDFTSKRIATGFIADPHLLQAPADDEOGRILDIYADV 497  
Db 421 AEFTVSGVATNIGFRLALREEDFTSKRIATGFIADPHLLQAPADDEOGRILDIYADV 480

QY 498 TVNPKHGVRPKDVAAPIDKLPNIKDLPLPRGSRDLKQLGPAAPARDLREODALAVTDTT 557  
Db 481 TVNPKHGVRPKDVAAPIDKLPNIKDLPLPRGSRDLKQLGPAAPARDLREODALAVTDTT 540

QY 538 FRDAHQSLLATRVRSFALKPAEAAVAKLTPELLSVEAWGGATYDVAMRFLFEDPWDRLE 617  
Db 541 FRDAHQSLLATRVRSFALKPAEAAVAKLTPELLSVEAWGGATYDVAMRFLFEDPWDRLE 600

QY 618 LREAMPNVNIQMLLRGNTVGYTPYPSVCRAFVKEAASGVVDIFRIFDALNDVYSQMRPA 677  
Db 601 LREAMPNVNIQMLLRGNTVGYTPYPSVCRAFVKEAASGVVDIFRIFDALNDVYSQMRPA 660

QY 678 IDAVLETNTAAVEAMAYSGDLSDPNEKLYLDIYLKMAEIVKSGAHILAIKDMAGLLR 737  
Db 661 IDAVLETNTAAVEAMAYSGDLSDPNEKLYLDIYLKMAEIVKSGAHILAIKDMAGLLR 720

QY 738 PAATVKLTALRRFDPPLVHVHTHTAGGQLATYFAAAQAGADAVDGCASAPLSTGTSQPS 797  
Db 721 PAATVKLTALRRFDPPLVHVHTHTAGGQLATYFAAAQAGADAVDGCASAPLSTGTSQPS 780

QY 798 LSAIVAAFAHTRRDTGUSLEAVSDELPYWEAVRGLYLPFESGTPGPTGRVRIEIPGQOL 857  
Db 781 LSAIVAAFAHTRRDTGUSLEAVSDELPYWEAVRGLYLPFESGTPGPTGRVRIEIPGQOL 840

QY 858 SNLRAQATFALGLADRFELIEDNYAAVNEMLRGPTKVTPTSSKVVGDLALHLVGAVDPADF 917  
Db 841 SNLRAQATFALGLADRFELIEDNYAAVNEMLRGPTKVTPTSSKVVGDLALHLVGAVDPADF 900

## RESULT 5

US-10-045-072-2  
; Sequence 2, Application US/10045072  
; Publication No. US20030027305A1  
; GENERAL INFORMATION:  
; APPLICANT: Sinskey, Anthony J.  
; APPLICANT: Lessard, Philip A.  
; APPLICANT: Willis, Laura B.  
; TITLE OF INVENTION: Pyruvate Carboxylase from Corynebacterium glutamicum  
; FILE REFERENCE: 1533.0790002  
; CURRENT APPLICATION NUMBER: US/10/045,072  
; CURRENT FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: US 09/677,575  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: US 09/220,081  
; PRIOR FILING DATE: 1998-12-23  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 2  
; LENGTH: 1140  
; TYPE: PRT  
; ORGANISM: Corynebacterium glutamicum  
US-10-045-072-2

Query Match 98.2%; Score 5759; DB 9; Length 1140;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1133; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 18 VSTHTSSTLPAPFKKILVANRGEIARAFRAALETGAATVAIYPREDRGSFHRSEAVR 77  
Db 1 MSTHTSSTLPAPFKKILVANRGEIARAFRAALETGAATVAIYPREDRGSFHRSEAVR 60

QY 78 IGTEGSPVKAYLIDIDEIIGAARKKADAIYPGYFLSENAQLARECAENGITFIGPTPEV 137  
Db 61 IGTEGSPVKAYLIDIDEIIGAARKKADAIYPGYFLSENAQLARECAENGITFIGPTPEV 120

QY 138 LDLTGDKSRANTAAKAGLPVLAESTPSKNIDDIIVKSAEGQTYPIFVKAVAGGGGRMRF 197  
Db 121 LDLTGDKSRANTAAKAGLPVLAESTPSKNIDDIIVKSAEGQTYPIFVKAVAGGGGRMRF 180

QY 198 VASPDRLKRLATEASREAAFGDGVYVERAVINPQHIIEVQILGDRGTGEVHHLYERDCS 257  
Db 181 VASPDRLKRLATEASREAAFGDGVYVERAVINPQHIIEVQILGDRGTGEVHHLYERDCS 240

QY 258 LQRRHQKVEIAPAOHLDPRLDRICADAVKFCRSIGYQAGTVEFLVDEKGNHVFTEMN 317  
Db 241 LQRRHQKVEIAPAOHLDPRLDRICADAVKFCRSIGYQAGTVEFLVDEKGNHVFTEMN 300

QY 318 PRIQVEHTVTEVTEDVLVKAQMRLAAGATLKELGLTQDKIKTHGAALOCRTITEDPNNG 377  
Db 301 PRIQVEHTVTEVTEDVLVKAQMRLAAGATLKELGLTQDKIKTHGAALOCRTITEDPNNG 360

QY 378 FRPDGTITAYRSPGAGVRLDGAOLGGEITAHFDSMLVKMTCRGSDFETAVARAQAL 437  
Db 361 FRPDGTITAYRSPGAGVRLDGAOLGGEITAHFDSMLVKMTCRGSDFETAVARAQAL 420

QY 438 AEFTVSGATNIGFLRALLREEDFTSKRIATGFIGDHPHLLQAPPADDEOGRILDLADV 497  
Db 421 AEFTVSGATNIGFLRALLREEDFTSKRIATGFIADHPHLLQAPPADDEOGRILDLADV 480  
QY 498 TVNKPCHVREPKVAAIDKLPNTKIDLPPLRGSRDLKQLGPAFAFARDLREQDALAVTDTT 557  
Db 481 TVNKPCHVREPKVAAIDKLPNTKIDLPPLRGSRDLKQLGPAFAFARDLREQDALAVTDTT 540  
QY 558 FRAHQSLATRVRSFALKPAEAAVAKLTPELLSVEAWGATYDVAMRFLFEDPWRLDE 617  
Db 541 FRAHQSLATRVRSFALKPAEAAVAKLTPELLSVEAWGATYDVAMRFLFEDPWRLDE 600  
QY 618 LREAMPNVNQMLLRGNTVGYTPYDPSVCRFAFVKEAASGVDFIRIFDALNDVQMRPA 677  
Db 601 LREAMPNVNQMLLRGNTVGYTPYDPSVCRFAFVKEAASGVDFIRIFDALNDVQMRPA 660  
QY 678 IDAVLETNTAAVEAMAYSGLSDPNKLYTLDYLLKMAEEIVKSGAHILAIKDMAGLLR 737  
Db 661 IDAVLETNTAAVEAMAYSGLSDPNKLYTLDYLLKMAEEIVKSGAHILAIKDMAGLLR 720  
QY 738 PAAVTKLVTALRREEDLPVHVHTHTAGGOLATYFAAAQAGADAVDGCASAPLSTTSOPS 797  
Db 721 PAAVTKLVTALRREEDLPVHVHTHTAGGOLATYFAAAQAGADAVDGCASAPLSTTSOPS 780  
QY 798 LSAIAAFAHTRRDTGLSLEAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGGQL 857  
Db 781 LSAIAAFAHTRRDTGLSLEAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGGQL 840  
QY 858 SNLRAQATLGLADREFELIEDNTAAVNEMLGRPTKVTTPSSKVGDLALHLVAGVDPADF 917  
Db 841 SNLRAQATLGLADREFELIEDNTAAVNEMLGRPTKVTTPSSKVGDLALHLVAGVDPADF 900  
QY 918 AADPOKYDIPDSVIAFLRGELGNPPCGWPEPLTRALEGRSEKAPLTEVPEEQAHLDA 977  
Db 901 AADPOKYDIPDSVIAFLRGELGNPPCGWPEPLTRALEGRSEKAPLTEVPEEQAHLDA 960  
QY 978 DSKERRNSNLRLFPKPTPEEFLEHRRFRFGNTSALDDREFYGLVEGRETLRLPDVVRT 1037  
Db 961 DSKERRNSNLRLFPKPTPEEFLEHRRFRFGNTSALDDREFYGLVEGRETLRLPDVVRT 1020  
QY 1038 LVLRLDAISPPDDKMRNVANVNGQIRPMVRDRDSVESVTATAEKADSSNKGHVAAPFA 1097  
Db 1021 LVLRLDAISPPDDKMRNVANVNGQIRPMVRDRDSVESVTATAEKADSSNKGHVAAPFA 1080  
QY 1098 GVVTVTVAEGDEVKAGDAVAIIEAMKMEATITASVDGKIERVVVPAATKVEGGDLIVVVS 1157  
Db 1081 GVVTVTVAEGDEVKAGDAVAIIEAMKMEATITASVDGKIERVVVPAATKVEGGDLIVVVS 1140

RESULT 6

US-09-815-242-5468  
; Sequence 5468, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; TITLE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA-011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: fastseq for Windows Version 4.0  
; SEQ ID NO 5468  
; LENGTH: 1147  
; TYPE: PRT  
; ORGANISM: Staphylococcus aureus  
US-09-815-242-5468

Query Match 42.4%; Score 2489; DB 10; Length 1147;  
Best Local Similarity 45.2%; Pred. No. 3.3e-160;  
Matches 527; Conservative 193; Mismatches 383; Indels 62; Gaps 14;

QY 30 KKILVANRGEIARAFRAALETGNAATVAIYPREDRGSFHRSEAFSEAVRIGTEGSPVKAYL 89  
Db 4 KKULVANRGEIARIFRAAAELDISTVAIYSNEKSLHRYKADESILVSGDLGPAESYL 63  
QY 90 DIDEIIAAGKVKADAIYPGYFLSENAQLARECAENGITFIGPTPEVLDLTGDKSAVT 149  
Db 64 NIERIDVAQANVDALHPGCGFLSENEQFARRCAEESIKFIGPHLEHLMFGDKVART 123  
QY 150 AAKGAGLPLV-ABSTPSKNIDDIKSAEGOTYPIFYKAVAGGGGRMRFVSSPDELRLKA 208  
Db 124 TAIKADLPVPGTDGPIKSYELAKEFAEAEAGFPLMIKATSGGGKGMRIREESELEDAF 183  
QY 209 TEASREAAAFGCGSVVVERAVINPOHIEVQILGDRTPGVVHLYERDCSLQRHOKVVEI 268  
Db 184 HRKSEAEKSGFSEVYIERIDNPKHIEVOVIGDEHGNIVHLPERDCSVORRHOKVVEV 243  
QY 269 APAQHLDPPELRDRICADAVKFCRSIGYGAGTVEFLV--DEKGNHVFIEHNPRIQVEHTV 326  
Db 244 APSVGLSPTLRQRICDAAIOLMENIKYVNAGTVEFLVSGDE--FFIEVNPRIQVEHTI 300  
QY 327 TEEVTEVDLVKQMRLAAGATL--KELGLTQDK--IKTHGAALQCRITTEDPNNFRPDTG 383  
Db 301 TEMVTGIDIVKTQITLVAAGADLFGEEINMPQOKDITTLGVAIQCRITTEDPNNFRPDTG 360  
QY 384 TIATYRSPGAGVRLD--GAAQLGGEITAHFDSMLVKTKCRGSDPETAVARAQRALAEFTV 442  
Db 361 TIATYRSPGAGVRLDAGDGFQGAETSPYDSSLVLLKTHAISPKQAEKMRVSLREMRI 420  
QY 443 SGVATNIGFLRALRREEDFTSKRIATGFIGDHPHLLQAPPADDEOGRILDLADVTKNP 502  
Db 421 RGVKTNIPFLINVMKKNKFTSGDYTKFIEETPELFDIQPSLDRGTCTLEYIGNVTIN-- 478  
QY 503 HGVKPKDVAAPIDKLPNIKDLPLP-----RGSRLDLKQLGPAAPAFARD 544  
Db 479 -----GFPNVKRPKPDYELASITPVSSSKIASFSGTKLLDEVGPKVAEW 525  
QY 545 LREQDALAVTDTTFRDAHQSLATRVRSFALKPAEAAVAKLTPELLSVEAWGATYDVAM 604  
Db 526 VKKQDDVLLDTTFRDAHQSLATRVRTKDMINIASATADVFKDGFSLMMGGGATFDVAY 585  
QY 605 RFLFEDPWRLDELREAMPNVNIQMLLRGNTVGYTPYDPSVCRFAFVKEAASGVDFIRI 664  
Db 586 NFLKENPWERLERLRKAIPNVLFOMLLRASNAVGYKNYPDNVHFKVQESAKAGIDVRI 645  
QY 665 FDALNDVSQMRPAIDAVLETNTAAVEAMAYSGLSDP--NEKLYTLDYLLKMAEEIVKSG 723  
Db 646 FDSLNVWDQMKVANEAVQEAQ--KISEGTICYTGDLINPERSNIYTYLEYVYVKLELEREG 704  
QY 724 AHILAIKDMAGLLRPAAVTKLVTALRREFDLPVHVHTHTAGGOLATYFAAAQAGADAV 783  
Db 705 FHILAIKDMAGLLKPKAAAYELIGELKAAVDLPILHLHTDTSNGLLTYKQADAGVDIID 764



```
Db 1058 SAVQVQKQAEPTNKEIGATMSGSLVQLVLRGDKVKEGQPLITEAMKMETTIEARFAG 1117
QY 1135 KIERYVVPAAATKVEGGDLIVV 1156
Db 1118 TVDHIYVEGEAIISSGDLLEEV 1139

RESULT 8
US-09-815-242-12361
; Sequence 12361, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELTRA-011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12361
; LENGTH: 1073
; TYPE: PRN
; ORGANISM: Staphylococcus aureus
US-09-815-242-12361

Query Match 39.2%; Score 2300; DB 10; Length 1073;
Best Local Similarity 44.8%; Pred. No. 1.9e-147;
Matches 489; Conservative 180; Mismatches 360; Indels 62; Gaps 14;

QY 104 DAIPGCGFSENAQLARCAENGITFIPTPEVLDTGDKSRVTAANKAGLPVL-AES 162
Db 2 DAIPGCGFSENEOPARCAEGKFIPTPEVLDTGDKSRVTAANKAGLPVLPGPD 61
QY 163 TPSKNIDIDIVKSAEGTPIFYKAVAGGGGMRVSSPDDELKRLKLAATESRAEAAFGD 222
Db 62 GPIKSVELAKEAEAGFLPMKATSGGGGKMRVREESELEDAFHRAKSAEAKSFGNS 121
QY 223 SVYVERAVINPOHIEVQILGDRGTGEVYVHLVYRDCSLQRRHQKVEITAPAHLDPELRDRI 282
Db 122 EVVIERYIDNPKHIEVQVIGDEHGNVHLVLRDCSVQRHQKVEVAPSVGLSPTLRQRI 181
QY 283 CADAVKFCRISGYQAGTVEFLV--DEKGNHVFIEIEMNPRIQVHTVTEVTEVDLVKAQM 340
Db 182 CDAATQLMENIKYVAGTVEFLVSGDE---FFFIENVRVQVHTITENMTGIDIVKTOI 238
QY 341 RLAAAGATL--KELGLTQDK-IKTHGAALOCRTITTPDNNRFRPDCTGTITAYRSPGAGVR 397
Db 239 LVAAGADLFEELINPQQKDIITGLVAIQCRITTEDPLNDFMPDGTITITAYRSSGGFGRV 298
QY 398 LD-GAQLGGEITAHFDSMLVKMTCRGSDFETAVARAQALAEFTVSGVATNIGFLRALL 456
```

```
Db 299 LDAGDFOGAIEISPYDLSLVKLSTHAISFKQAEERKVRSLRMRIRGVKTNIPFLINVM 358
QY 457 REEDFTSKRIATCFICDHPHILLOAPPADDEOGRILIDYADVTYVKNKPHGVPRKDVAPIDK 516
Db 359 KKKFTSGDYTTTKFIEETPELFDIQPSLDGRGKTLEYIGNVTIN-----G 403
QY 517 LPNIKDLPLP-----RGSRDRLKQLGPAAPFARDLREQDALAVTDITTF 558
Db 404 FNVEKRPKPDYELASIPTVSSSKIASFSGTQKOLLDEVPKGVAEVWKKODDVLITDTTF 463
QY 559 RDAHQSLLATRVRSFALKPAEAVAKLTPELLSVEAWGATYDVAMRFTEFEDPMDRLDEL 618
Db 464 RDAHQSLLATRVRTKDMINIASKTADVFKDGFSELMGAGTADVAYNFLKENPWERLERL 523
QY 619 REAMPNVNIOMLLRGNTVGYTPYDSCVACRAFKVKEAASGVDIFRIFDALNDVSOHRPAI 678
Db 524 RKAIPNVLFQMLLRASNAVGYKNYPDNIHKFVQESAKAGIDVFRIFDLSLWYDQMKVAN 583
QY 679 DAVLETNTAVAEVAMAYSGDLSDP-NEKLYTLDYILKMAEIEVKSGAHLIAIKDMAGLLR 737
Db 584 EAVQEAG-KISECTICYTGDILNPERSNIYLEYVYVKLAKELEREGFHILAIKDMAGLLK 642
QY 738 PAAVTKLVLTALRREFDLPVHVHTHDAGQLATYFAAAQAGADAVDAGSAPISGTTSQPS 797
Db 643 PRAAYELIGELKAAVDLPILHLHTDTSNGLLTYKQADAGVDIIDTAVASMSGLTSQPS 702
QY 798 LSAIYAFAHTRDRTGLSLEAVSDLEPYWEAVRGLLYLPESGTPGTPGRVYRHEIPGGQL 857
Db 703 ANSLYALNGFPRHLRTDIEGMESLSHYSTVRTYYSDFESDIKSPNTEIYQHMPGGQY 762
QY 858 SNLRQAATAGLADRPEDIEDNYAAVNEMLGRPTKVTPTSSKVVDLALHLVAGVDPADF 917
Db 763 SNLSQAQSLGLGERDEVDKMYRRVNFLEGDIVKVTPTSSKVVDGMLALYVQNDLDEQSV 822
QY 918 AADPOKYDIPDSVIAFLRGLGNPPGGWPEPLRTRALEGRSEKAGPLTEVPEEQAHLDA 977
Db 823 ITDGYKLDFFESVVSFKEIGOPVNGKDLQAVILKQE-----ALTARPGEYLEPVD 878
QY 978 DDSKE-----RRNSLRNLFLPKTEFELEHRRRFGNTSALDDREFFYGLVEGRE 1026
Db 879 EKVERLEEEQOGPVTEQDIISYVLYPKYVEQYIOTRNYGNLSLDDTPTFFFGMRNET 938
QY 1027 TLIRLPDVRTPLLRDLAISEDDKGMNVVANNVNGIOPMRVDRSRVESVTATAEKADS 1086
Db 939 VEIEI-DKGRLLIKLETISEPDENGNRTIYYAMNQOARRIYIKDENVHTNANVKPADK 997
QY 1087 SNKGHVAAFPAGVVT-VTVAGDEVKAGDAVAIIEMKMEATITASVDGKIERYVVPAAAT 1145
Db 998 SNPSHIGAQMPSGVTEVKVSVGETVKANQPLLITEAMKMETTIQAPFDGVIKQVTVNNGD 1057
QY 1145 KVEGGDLIVV 1156
Db 1058 TIATGDLIIIEI 1068

RESULT 9
US-09-767-479-6
; Sequence 6, Application US/09767479
; Patent No. US20010036654A1
; GENERAL INFORMATION:
; APPLICANT: Haselkorn, Robert
; Gornicki, Piotr
; TITLE OF INVENTION: ACETYL-CoA CARBOXYLASE COMPOSITIONS AND
; METHODS OF USE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/767,479
  FILING DATE: 22-Jan-2001
  CLASSIFICATION: Unknown
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/468,793
  FILING DATE: <Unknown>
  APPLICATION NUMBER: US SN 07/956,700
  FILING DATE: 02-OCT-1992
  APPLICATION NUMBER: PCT/US93/09340
  FILING DATE: 30-SEP-1993
ATTORNEY/AGENT INFORMATION:
  NAME: Kitchell, Barbara S.
  REGISTRATION NUMBER: 33,928
  REFERENCE/DOCKET NUMBER: ARCD:152/KIT
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (512) 418-3000
  TELEFAX: (713) 789-2679
  TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
  LENGTH: 447 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-767-479-6

Query Match      17.6%  Score 1032.5; DB 10; Length 447;
Best Local Similarity 48.2%; Pred. No. 3.6e-62;
Matches 216; Conservative 71; Mismatches 148; Indels 13; Gaps 4;

QY 29 FKKILVANRGETIAVRAALETGAATVAIYPREDRGSEFHSFASFAVRIGTGPSPVKAY 88
DB 3 FDKILLANRGEIALRLRACEENGATIAVHSTVDNRNALHVQLADEAVCIG-EPASAKSY 61

QY 89 LDIDEILGAAKVKADAIYPGYGFLSENQAQARECAENGITFTGPTPEVLDLTGDKSRV 148
DB 62 LNIPLNIAALTRNASHIPGYGFLSENQAQARECAENGITFTGPTPEVLDLTGDKSRV 121

QY 149 TAARKAGLPVLAES-----TPSKNIDDIVKSAEGQTYPIFVKAVAGGGGRGMRVSSPDE 203
DB 122 ETMQKAGVTPVPGSEGLVETEGLE---LAKDIGYPMVKATAGGGGRGMRVSSPDE 177

QY 204 LRKLATEASREAAAFGDSGVYVERAVINPOHIEVQILGDRGTGEVHLYERDCSLQRRHQ 263
DB 178 FVKFLAAGGAGAGAGNAGYIEKFTERPRIEFQILADNYGNVHILGERDCSIQRRNQ 237

QY 264 KVVETAPQAHLDPDLRDRICADAVKFCRSIGYOGAGTVEFLVDEKGNHVFIEKNPRIQVE 323
DB 238 KLLBEAPSPALDRLREKMGQAQVAAQAFINYTGAGTIEFLDRSGQFYEMNTRIQVE 297

QY 374 HTVTEEVTEVDLVKAQMLAAGATLKLGLTODKIKTHGAALOCRTTDPNNGFRPDTC 393
DB 298 HPVTEWTVGVDLVEQIRIAGGERLR---LTQDQVLRGHAIECRINAEDPDHDFRPAQ 354

QY 384 TITAYRPGGAGVRLDGAALGGBITAHDFTSMVKKMTCRSDSETAVARAQRLAETVS 443
DB 355 RISGLPPGPGVGRIDSHVYTDVQIIPYDLSLIGKLVWGPDRATAINRMKRALRECAIT 414

QY 444 GVATNIGFLRALLREEDFTSKRIATGFI 471
DB 415 GLPTTIGFHORIMENPQFLQGNVSTSFV 442

RESULT 10
US-09-767-479-8

```

```

Sequence 8, Application US/09/767479
Patent No. US20010036654A1
GENERAL INFORMATION:
  APPLICANT: Haselkorn, Robert
  Gornicki, Piotr
  TITLE OF INVENTION: ACETYL-CoA CARBOXYLASE COMPOSITIONS AND METHODS OF USE
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
  ADDRESSER: Arnold, White & Durkee
  STREET: P. O. Box 4433
  CITY: Houston
  STATE: Texas
  COUNTRY: United States of America
  ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/767,479
  FILING DATE: 22-Jan-2001
  CLASSIFICATION: Unknown
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/468,793
  FILING DATE: <Unknown>
  APPLICATION NUMBER: US SN 07/956,700
  FILING DATE: 02-OCT-1992
  APPLICATION NUMBER: PCT/US93/09340
  FILING DATE: 30-SEP-1993
ATTORNEY/AGENT INFORMATION:
  NAME: Kitchell, Barbara S.
  REGISTRATION NUMBER: 33,928
  REFERENCE/DOCKET NUMBER: ARCD:152/KIT
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (512) 418-3000
  TELEFAX: (713) 789-2679
  TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
  LENGTH: 453 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-767-479-8

Query Match      17.1%  Score 1005.5; DB 10; Length 453;
Best Local Similarity 47.1%; Pred. No. 2.5e-60;
Matches 209; Conservative 72; Mismatches 158; Indels 5; Gaps 3;

QY 29 FKKILVANRGETIAVRAALETGAATVAIYPREDRGSEFHSFASFAVRIGTGPSPVKAY 88
DB 3 FKKILLANRGEIALRLRACEENGATIAVHSTVDNRNALHVQLADEAVCIG-EAASSKSY 61

QY 89 LDIDEILGAAKVKADAIYPGYGFLSENQAQARECAENGITFTGPTPEVLDLTGDKSRV 148
DB 62 LNIPLNIAALTRNASHIPGYGFLSENQAQARECAENGITFTGPTPEVLDLTGDKSRV 121

QY 149 TAARKAGLPVLAESTP-SKNIDDIVKSAEGQTYPIFVKAVAGGGGRGMRVSSPDELRLK 207
DB 122 ETMQKAGVTPVPGSEGLTVDVSAKVAEIGYPMVKATAGGGGRGMRVREPDLLEKL 181

QY 208 ATEASREAAAFGDSGVYVERAVINPOHIEVQILGDRGTGEVHLYERDCSLQRRHQVVE 267
DB 182 FLAAGCEAAAFGNPGLYLEKFDPRHVEFOILADAYGNVHVLGERDCSIQRRHOKLLE 241

QY 268 IAPAQHLDPDLRDRICADAVKFCRSIGYOGAGTVEFLVDEKGNHVFIEKNPRIQVEHPT 327
DB 242 EAPSPALSADLRQKMGDAVKVAQAIGYIGACTVEFLVDATGDNFYEMNTRIQVEHPT 301

```

```
QY 328 EEVTEVDLVKAQMRILAAGATLKELGLTQDKIKTHGAALQCRITTEDPNNNGFRPDGTGTTA 387
Db 302 EMITGLDLIAEQIRIAQGAELR---FROADIQLRHAIECRINAEDPEYFNFRNPGRIIG 358
QY 388 YRSPGAGVRLDGAALGGEITAHFDSMLVKMTCRGSDFETAVARAQALAEFTVSGVAT 447
Db 359 YLPPGPGVGVDSHVYTDYIEPPYDSLIGLKLIVMGATREELARMORALRECAITGLPT 418
QY 448 NIGCFLLALLREEDFTSKRIATGFI 471
Db 419 TLSFHQLMLQMPFELRGELYTNFV 442

RESULT 11
US-09-815-242-5215
; Sequence 5215, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5215
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5215

Query Match 16.7%; Score 979.5; DB 10; Length 471;
Best Local Similarity 45.9%; Pred. No. 1.5e-58;
Matches 209; Conservative 74; Mismatches 157; Indels 15; Gaps 6;
```

```
QY 30 KKLIVANRGEIAVRAFRALLETGAATVAIYPREDRGSFHRSEAFSEAVRIGTEGSPVKAYL 89
Db 3 KKLIVANRGEIAVRIRACAEIGRSVAVYSEADRHVHKRADEAHSIGAD--PLAGYL 60
QY 90 DIDEIIGAARKVKADATYPGYFGLSENAQARECAENGITFTGPTPEVLDLTDGDKSRVAT 149
Db 61 NPRALVNLAVERGSCDALHPGYFGLSENAELALCAERGIKFGPSAQVIRMGDKTEAR 120
QY 150 AAKKAGLPVLAESTSK--NIDDI---VKSAGQOTYPIPVKAVAGGGGRMFRVSSPDEL 204
Db 121 SMIAAGVP---CTPGTEGNVADLAELAREERIGYPMVKATSGGGGGRIRRCNSREEL 176
QY 205 RKLATEASREAAAGDCGVYVERAVINPOHTEVQILGDRTEGVHLYERDCSLQRHOK 264
Db 177 EQAFPRVISEATKAFGSAEVEFKECIVNPKHIEAQILADSFNGTVHLFERDCSIQRNOK 236
```

```
QY 265 VVEIAPAQHLDPDLRDRICADAVKFCRSIGYOGAGTVEFLVDEKGNHVFTEMNPRIQVEH 324
Db 237 LIEIAPSPOLTPQRAYIGDLAVRAAKAVGYNAGTVEFLAD-GEVYFEMMTRVQVEH 295
QY 325 TVTEEVTEVDLVKAQMRILAAGATLKELGLTQDKIKTHGAALQCRITTEDPNNNGFRPDGTG 384
Db 296 TTEETITGDIVVREQIRIASGL---ELSVKQDDIVHRGYALQPRINAEDPKNNFLPSFGK 352
QY 385 ITAYRSPGAGVRLDGAALGGEITAHFDSMLVKMTCRGSDFETAVARAQALAEFTVSG 444
Db 353 ITRYAPGPGVGTDTAIYTGTYTIPYDSMCKLKLIVALTWEEALDRGLRDLDMRVQG 412
QY 445 VATNIGFLRALLREEDFTSKRIATGFIGDHPHLLQ 479
Db 413 VKTTAPYQEIILRNPFERSGQNTSFVESHPELTQ 447

RESULT 12
US-09-815-242-13885
; Sequence 13885, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13885
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13885

Query Match 16.3%; Score 956; DB 10; Length 449;
Best Local Similarity 47.6%; Pred. No. 5.5e-57;
Matches 206; Conservative 61; Mismatches 158; Indels 8; Gaps 5;
```

```
QY 31 KKLIVANRGEIAVRAFRALLETGAATVAIYPREDRGSFHRSEAFSEAVRIGTEGSPVKAYL 90
Db 4 KLVIANRGEIALRILRACKELGKTVAVHSSADRLKHVLLADETCVIGPAPS-VKSYLN 62
QY 91 IDEITGAARKVKADATYPGYFGLSENAQARECAENGITFTGPTPEVLDLTDGDKSRVAT 150
Db 63 IFAISAAEITGAVAIHPGYFGLSENAFAEQVERSGFIFIGPKADTIRLMGDKVSAITA 122
QY 151 AKKAGLPVLAESTSKNIDDIKVSAGQT--YPIFVKAVAGGGGRMFRVSSPDELRLKLA 208
Db 123 MKKAGTYTPVGSDDPLGDDMNANRAHAKRIGYPVIIKASGGGGGRMFRVRSDAELAQSI 182
```





Qy	29	FKKILVANRGEIAVRAFAALETGAATVAIYIPREDRGSFHRSEFASEAVRIGTEGSPVKAY	88
Db	2	FKKILVANRGEIAVRIIRAARELGIAVAVYSTADKEALHTLLADEAVCIG-PGKATESY	60
Qy	89	LDIDEIITCAAKVKADAIYPCYGLFSENAOLARECAENGITFGPPEVLIDLTDGKSRAV	148
Db	61	LNINAVLSAAVLTEAEAIHPGFLSNSFAFMCEVGFKFGPSGHVMDMGDKINAR	120
Qy	149	TAACKAGLPVLAESTPS-KNIDDIIVKSAEGTQYIPFKAVAGGSGRGMREVSPPDELRLK	207
Db	121	AQMIKAGVPVPGSDGEVHNSEEAALIVAEKIGYPMVLKASAGGGKGRKVEKPPDLVSA	180
Qy	208	ATEASRAEAAFGDGYVVERAVINPOHIEVQILGORTGEVHLYBRDCSLQRHOKVVE	267
Db	181	FETASSEKANYGNAMYIERVYPARHIEVQILGDEGHVHIILGDERCSLORNNOKVLE	240
Qy	268	IAPAQHLDPELDRICADAVKFCRSIGYQAGAGTVEFLVDE-KGNHVFIEMNPRIOQVHTV	326
Db	241	ESPSIAIGTKLLRHEIGAATAVAAEEFVGYENAGTIEFLDDEASSNFYEMMTRVQVEHPV	300
Qy	327	TEEVTEVDLVKAMRLAAGATLKEGLGTQDKIKYTHGAALOCRTITTEDPNNGFPDGTGIT	386
Db	301	TEFVSGVDIVKEQICIAAG--QPLSVYQBEDIVLRGHAIECRTNAENAPAFNEAFSPGKIT	357
Qy	387	-AYRSPCGAVRLDGAOLGGETTAHFDSMLVKWTCRGSDFEFAVARAQALAEFTVSGV	445
Db	358	NLYLPSGCVGLRVDSAVYPGYTTPPYDYSMAKILYHGENRFPDLMKMQBALYELIEGV	417
Qy	446	ATNITFLALLREEDFTSKRIATGFI	471
Db	418	QTNADFOLDLISDRNVIAGDYDTSFL	443

; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari L.  
 ; APPLICANT: Zyskind, Judith W.

```

: : TITLE OF INVENTION: Identification of Essential Genes in
: : TITLE OF INVENTION: Prokaryotes
: : FILE REFERENCE: ELITRA.011A
: : CURRENT APPLICATION NUMBER: US/09/815,242
: : CURRENT FILING DATE: 2001-03-21
: : PRIOR APPLICATION NUMBER: 60/191,078
: : PRIOR FILING DATE: 2000-03-21
: : PRIOR APPLICATION NUMBER: 60/206,848
: : PRIOR FILING DATE: 2000-05-23
: : PRIOR APPLICATION NUMBER: 60/207,727
: : PRIOR FILING DATE: 2000-05-26
: : PRIOR APPLICATION NUMBER: 60/242,578
: : PRIOR FILING DATE: 2000-10-23
: : PRIOR APPLICATION NUMBER: 60/253,625
: : PRIOR FILING DATE: 2000-11-27
: : PRIOR APPLICATION NUMBER: 60/257,931
: : PRIOR FILING DATE: 2000-12-22
: : PRIOR APPLICATION NUMBER: 60/269,308
: : PRIOR FILING DATE: 2001-02-16
: : NUMBER OF SEQ ID NOS: 14110
: : SOFTWARE: FastSEQ for Windows Version 4.0
: : SEQ ID NO 13617
: : LENGTH: 455
: : TYPE: PRT
: : ORGANISM: Streptococcus pneumoniae
: : US-09-815-242-13617

```

Query Match	16.0%	Score 936.5;	DB 10;	Length 455;
Best Local Similarity	45.7%	Pred. NO. 1.2e-55;		
Matches 204; Conservative	76;	Mismatches 159;	Indels 7;	Gaps 5;



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 18:58:45 ; Search time 54 Seconds  
(without alignments)  
2059.769 Million cell updates/sec

Title: US-09-974-973-2  
Perfect score: 5865  
Sequence: 1 MTATLGGLLKIGITLVST.....RVVVPAAATKVEGGDLIVVVS 1157

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_73:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3713.5	63.3	1127	D70671	pyruvate carboxyla
2	2550.5	43.5	1178	A47255	pyruvate carboxyla
3	2548.5	43.5	1178	JC4391	pyruvate carboxyla
4	2535.5	43.2	1150	A83978	pyruvate carboxyla
5	2530	43.1	1148	F69685	pyruvate carboxyla
6	2528.5	43.1	1178	JC2460	pyruvate carboxyla
7	2494	42.5	1146	AH1208	pyruvate carboxyla
8	2491	42.5	1146	AC1565	pyruvate carboxyla
9	2490.5	42.5	1144	D37227	pyruvate carboxyla
10	2489	42.4	1150	D89881	pyruvate carboxyla
11	2474.5	42.2	1174	C37686	pyruvate carboxyla
12	2474.5	42.2	1174	A82111	pyruvate carboxyla
13	2465.5	42.0	1158	A83285	pyruvate carboxyla
14	2461	42.0	1175	T20346	pyruvate carboxyla
15	2457	41.9	1178	Q7BYFP	pyruvate carboxyla
16	2447	41.7	1185	T39734	pyruvate carboxyla
17	2436	41.5	1180	S46094	pyruvate carboxyla
18	2417.5	41.2	1195	T43735	pyruvate carboxyla
19	2413	41.1	1137	E86708	pyruvate carboxyla
20	2119.5	36.1	984	T44608	pyruvate carboxyla
21	1079	18.4	501	D64453	biotin carboxylase
22	1071.5	18.3	477	G70427	biotin carboxylase
23	1040	17.7	472	A70432	biotin carboxylase
24	1032.5	17.6	447	A53311	biotin carboxylase
25	1032.5	17.6	447	AH1923	biotin carboxylase
26	1020	17.4	506	D69277	biotin carboxylase
27	996	17.0	491	A69123	biotin carboxylase
28	979.5	16.7	471	G82966	probable biotin ca
29	978.5	16.7	448	S74380	biotin carboxylase

ALIGNMENTS

RESULT 1

D70671

pyruvate carboxylase (EC 6.4.1.1) - Mycobacterium tuberculosis (strain H37RV)  
C:Species: Mycobacterium tuberculosis  
C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 11-Jan-2002  
C:Accession: D70671; S73055  
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: D70671

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1127 <COL>  
A:Cross-references: GB:283018; GB:AL123456; NID:g3261671; PIDN:CAB05410.1; PID:g169.8  
A:Experimental source: strain H37RV  
R:Smith, D.R.; Robison, K.  
submitted to the EMBL Data Library, September 1994  
A:Description: Mycobacterium tuberculosis cosmid tbc2.  
A:Reference number: S73053

A:Accession: S73055

A:Molecule type: DNA

A:Residues: 1-353, TRAGSARCDPPAVPVSAWTAAPTWRNQPVLRLHAGQADLS',396-1115,'EWRAETCWW'

A:Cross-references: EMBL:000024; NID:g560506; PIDN:AAA50948.1; PID:g560527

C:Genetics:

A:Gene: pca; pyc

A:Start codon: GTG

C:Superfamily: pyruvate carboxylase; biotin carboxylase homology; lipoyl/biotin-binc

C:Keywords: biotin binding; ligase; mitochondrion

F:4-457/Domain: biotin carboxylase homology <BCH>

F:1055-1127/Domain: lipoyl/biotin-binding homology <LPB>

F:1093/Binding site: biotin (Lys) (covalent) #status predicted

Query Match 63.3%; Score 3713.5; DB 2; Length 1127;

Best Local Similarity 64.4%; Pred. No. 2.6e-196;

Matches 730; Conservative 153; Mismatches 239; Indels 11; Gaps 6;

Qy 29 FKKILVANGETAVRAFALETGTATVAIYPREDGSGFHSFASBAVRITGEGSPVKAY 88

Db 2 FSKVLVANGETAIFRAFRAYELGVCTVAVPYEDRNSQHLKADSEYQIGDIGHPVHAY 61

Qy 89 LDIDEITGAKKVADAIYPGYGFLSENAQLARECAENGITFTGTPPEVLDTGTGDKSRV 148

Db 62 LSVDEIVTARRAGDAIYPGYGFLSENPDLAAACAAGISFVGSFAEVLKAGNSRAI 121

Qy 149 TAAKAGLPLVAESTPSKNIIDIVKSAEGQTPYIFKAVAGGGGRCMRVFSFDELRLKLA 208

Db 122 AAAREAGLPLVMSAPSASVDELLSVAGGMPPLFLFKAVAGGGGRCMRVFSFDELRLKLA 181

QY 209 TEASREAAFGDGVVVERAVINPQIHVEIQLDRTGEVHLYERDCSLQRRHQKVEI 268  
DB 182 EASREAESFGDVTYVLEQAVINPRHIEVQILADNLGDLVHLYERDCSVORRHQKVEL 241  
QY 269 APAOHLDELPCADAKVFCRSIGCGAGTVEFLVDEKGNHVFIEIMNPRIOVEHVTI 328  
DB 242 APAHLDAELRYKMCVDVAFARIHIGYSCAGTVEFLDERGEYFIEIMNPRVQVEHVTI 301  
QY 329 EVTEVDLVKAQMLRAAGATLKEGLTQDKIKTHGAALOCRTITDPNNGRPDGTITAY 398  
DB 302 EITVDLVASQIRTAAGTLEQLGREDIAPHGAALOCRTITDPNNGRPDGTITAY 361  
QY 399 RSPGAGVRLDGAALGGEITAHFDSMLVMTKRGSDFTETAVARAQALAEFTVSGVATN 448  
DB 362 RTAGAGVRLDGSYNLGAETSPYFDSMLVMTKRGSDFTETAVARAQALAEFTVSGVATN 421  
QY 449 IGFRLALREEDFTSKRIATGFIGDHPHLLQAPPADDEQGRILDYLDVTVNPKPHGVPRK 508  
DB 422 IPFLQAVLDLDPFRAGRTVTSFIDERPQLLTARASADRGTKILNFLADVTVNPNYGRSP 481  
QY 509 DVAAPIDKLPNIKDLPL---PRGSRDLKOLGPAAPAFAROLREODALAVDTTFRDAHQ 564  
DB 482 TI-YPDOKLP---DLDLRAAPPAGSKORLVKLGPEGFARWLRESAAVGVDTTFRDAHQ 537  
QY 565 LLATRVRSFALKPAEAAVAKLTPELLSVEMAGGATYDVAMRFLFEDPDWRDLDELREAMPN 624  
DB 538 LLATRVRTSGLSRVAPVARTMPQLLSVECGGATYVALRFLKEDPWERLATLRAAMPN 597  
QY 625 VNIQMLLRGNVTGTYPPDSVCFRAFYKEAASCGVDIFRIFDALNDVYSQMRPAIDAVLET 684  
DB 598 ICLQMLLRGNVTGTYPPPEIVTSFAFQEAATATCIDIFRIFDALNNTESMRPAIDAVRET 657  
QY 685 NTAVAEVAMVSGDLSDPNEKLYTLDYLYKMAEIVKSGAHILAIKDMAGLIRPAATVKL 744  
DB 658 GSATAEVAMVCTGDLTDGEOQLYTLDYLYKLAEOIVDAGARVLAIKDMAGLIRPAATVKL 717  
QY 745 VTLARRDELPHVHTHTAGCOLATYFAAAQACADAVDGCASAPLSGTTSPOLSATVAA 804  
DB 718 VSALRSRDELPHVHTHTDPGQLASVVAWHAGADAVDGAAPAGTTSOPALSSVAA 777  
QY 805 FAHTRDRLSLEAVSDLEPWEAVRGYLPFESGTPGTPGVRYRHEIPGQSLNLRQA 864  
DB 778 AAHTFYDGLSLSAVCALEPWEALRYVAFPSGLPGTPGVRYRHEIPGQSLNLRQA 837  
QY 865 TALGLADREFELIDYNAVNEMLGRPTKVTPESSKVVGDLALHLVAGVDPADFADPQKY 924  
DB 838 TALGLDRFEIEEAYAGADRVGLRVKVTPTSKVVGDLALALVAGVSADEFASDPARF 897  
QY 925 DIPDSVIAFLRGELGNPPGWPPELRTALRGSEKAPLTPVEPEEQAHLDADDSKERR 984  
DB 898 GIPESVGLRGELGDPGWPPELRTAALAGGAAR-PTAQLAADDEIALSSVGAK-RQ 955  
QY 985 NSLRLFLPKPTEFFLEHRRRFGNTSALDDREFFYGLVEGRETLIRLPDVRTPLLVRLDA 1044  
DB 956 ATLRLFLPPTKEFNEHREAYGDTLSLSANQFFYGLRGEEHVRKL-ERGVELLIGLEA 1014  
QY 1045 ISEPDKGMNVVANVNGOIRPMVRDRSPVESVTATAEKADSSKNKHVAAPFAGVVTVTY 1104  
DB 1015 ISEPDGCMRVIMLNGQLRPLVLRDSIASVPAEAKADRGNGPHIAFPAGVVTIGV 1074  
QY 1105 AEGDVKKADAVAIETAMKMEATTASVDGKIERYVVPATKVEGGDLVVVS 1157  
DB 1075 CVGERVAGQTIATIEAMKNEAPITAPVAGTVVAVSDTAQVEGGDLVVVS 1127

RESULT 2  
A47255  
pyruvate carboxylase (EC 6.4.1.1) precursor [similarity] - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 01-Feb-2002  
C:Accession: A47255  
R:Zhang, J.; Xia, W.L.; Brew, K.; Ahmad, F.  
Proc. Natl. Acad. Sci. U.S.A. 90, 1766-1770, 1993

A:Title: Adipose pyruvate carboxylase: amino acid sequence and domain structure deduc  
A:Reference number: A47255; MUID:93189578; PMID:8446588  
A:Accession: A47255  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1178 <ZNA>  
A:Experimental source: GB:109192; MID:g293743; PIDN:AAA39737.1; PID:g293744  
A:Cross-references: GB:109192; MID:g293743; PIDN:AAA39737.1; PID:g293744  
A:Note: sequence extracted from NCBI backbone (NCBIN:126874, NCBIPI:126875)  
C:Superfamily: pyruvate carboxylase; biotin carboxylase homology; lipoyl/biotin-bindi  
C:Keywords: biotin binding; ligase; mitochondrion  
F:1-20/domain: transit peptide (mitochondrion) #status predicted <TRP>  
F:21-1178/product: pyruvate carboxylase #status predicted <NAT>  
F:39-494/domain: biotin carboxylase homology <BCH>  
F:1105-1178/domain: lipoyl/biotin-binding homology <LPB>  
F:1144/Binding site: biotin (Lys) (covalent) #status predicted

Query Match 43.5%; Score 2550.5; DB 1; Length 1178;  
Best Local Similarity 46.1%; Pred. No. 2.8e-132;  
Matches 543; Conservative 192; Mismatches 407; Indels 35; Gaps 13;

QY 7 GGLLKGIITLVSTHTSSTLPA-----PKKILVANRGETAVRAFRAALETGAA 54  
DB 9 GGLRLGV-----RRSSAPVAPSNVRRLKYKPKKVMVANRGEIATVFRACTELGR 62  
QY 55 TVAIYPRDRGSGFHSRFASEAVRIGTEGSPVKAYLIDIEIIIGAAKKVKADAIYPGVGFLS 114  
DB 63 TVAVYSEQDTGQMHRQKADAEVYLIGRLAPVQAYLHPDIIVKAKENGVDVAPGVGFLS 122  
QY 115 ENAQLARECAENGTIFGPTPEVLDLTGDKSRVATAAKKAGLPVL-AESTPSPKNTDDIVK 173  
DB 123 ERADFQAQODAGVRFIGPSPVVRKMGDKVEARAIAAAGVPVPGTDSPISSLHEAHE 182  
QY 174 SABGQTYPIFKVAVAGGGGRMFVSSDPDLRKLKATEASREAFAGDGVVVERAVINP 233  
DB 183 FSNTPGFPFIKAAVGGGGRGVVHSEYELEENYTRAYSEALAAFGNAGLVEKFIKP 242  
QY 234 OHLEVOILGDRTEGVVHLYERDCSLORRHQKVETAPAOHLDELPCADAVAFCSR 293  
DB 243 RHVEVOILGDOYNTUHLRYERDCSIORRHQKVETAPATHLPQLSRUTSDSVKLAKOV 302  
QY 294 GYOGACTVFLVDEKGNHVFIEIMNPRIOVEHVTVEVDLVKAQMLRAAGATLKEGL 353  
DB 303 GYENACTVFLVDEKGNHVFIEIMNPRIOVEHVTVEVDLVKAQMLRAAGATLKEGL 362  
QY 354 TQDKIKTHGAALOCRTITDPNNGRPDGTITAYRSPGGAGVRLDGAAL-EGEITAHF 412  
DB 363 RQENIRINGCATQCRVTTEDPARSFQDPTGRIEVRFSRSGEGMGIRLDNASAFQCAVISPHY 422  
QY 413 DMLVMTKRGSDFTETAVARAQALAEFTVSGVATNIGFLRALLREEDFTSKRIATGFIG 472  
DB 423 DSLLVKVIAGKHDPHTAATKMSRALAEFRVGVKTNIPFLQVNNQQLAGTVDQTFID 482  
QY 473 DHPHLLQAPPADDEQGRILDYLDVTVNPKPHGVPRKQDA-APID-KLPNIKDLPGRGSR 530  
DB 483 ENPELFQLRPAQNRACKLHLYGHVNVNGTTPIPVNVSPSPVDPAPVVPVPGPPAGFR 542  
QY 531 DRLKOLGPAARFLEQDALAVDTTFRDAHQSLATVRSFALKPAEAAVAKLTPELL 590  
DB 543 DILLREGPGFAKAVRNHOGCLLMDITFRDAHQSLATVRSFALKPAEAAVAKLTPELL 602  
QY 591 SVEAWGATYDVAMRFLFEDPDWRDLDELREAMPNVIQMLLRGNVTGTPYDPSVCFRA 650  
DB 603 SMENWGATYDVAMRFLFECPPWRRLQELRELIPNIPFQMLLRGANAVGTYNPDNVVFKF 662  
QY 651 VKEAASSGVDFIRIFDALNDVYSQMRPAIDAVLETINTAVAEVAMVSGDLSDPNEKLYTLD 710  
DB 663 CEVAKENGMDVFRVFDLSNLYLPNMLLGMGAAGSAG-GVVEAAISYTGVDVADPSRTKSYLE 721  
QY 711 YLLKMAEEIVKSGAHILAIKDMAGLIRPAATVKLTALRREF-DLPVHVHTHTDTAGGOLA 769  
DB 722 YVMGLAEELVRAGTHILCISKDMAGLIRPAATVKLTALRREF-DLPVHVHTHTDTAGGOLA 781

QY 770 TYFAAQAQADAVDASAPLSTGTTSPSLSAIVAFAHTRDTGSLSEAVSDLEPYWEAV 829  
Db 782 AMLACAQAGADVVDVDSMGSTSPSMGALVACTKGTPLDTEPLERVDSEYWECA 841  
QY 830 RGLYLPFESTGPTGTR--VYRHEIPGGQLSNLRAQATAGLADRELLIEDNYAANVEML 887  
Db 842 RGLYAAFDCTATMKSGNSDVSNEYENIPGGQYTNLHFQAHSMGLSGSKFKEVKAYVEANQML 901  
QY 888 GRPTKVTSPSKVGVGLALHLVAGVDPADFAADPOKYDIPDSVIAFLRGLGNPPGWPE 947  
Db 902 GDLIKVTSPSKVIGDLAQFVWQNGLSRAEAEAEELSPPRSVVEFLOQYIGIPHGFFPE 961  
QY 948 PLRTRAL-----EGRSEKGAPLTEVPEEEQAHLDA-DDSKERRNSLNRLLPKPTPEEL 1000  
Db 962 PFRSKVLKDLPRIEGRPGASLPPLNLKELEKDLIDRHGEVTPEDVLSAAMPDVFAQFK 1021  
QY 1001 EHRRRFGNTSALDDREFFVGLVEGRETLLRLPDVTRPLVRLDAISEPDDKGRNVVAV 1060  
Db 1022 DFTATFGPLDSLNRLFLQGPKEAEFEVELEGRGT-LHIKALAVSDLNRAQGRQVFFEL 1080  
QY 1061 NGQIRPMVRDRSVESVTATAEKADSSNKGHVAAFPAG-VVTVTVAEGDEVKAGDAVATI 1119  
Db 1081 NQOLRSILVKDQAMKEMHFHFKALKDKVKGQIGAPMPGKVIDIKVNAAGDKVAKGQPLCVL 1140  
QY 1120 EAMKMEATITASVGGKIERVVVPAATKVEGGDLVVV 1156  
Db 1141 SAMKMETVVTSPMEGTIRKVVHTKDMTLEGDDLILEI 1177

RESULT 3  
JC4391  
pyruvate carboxylase (EC 6.4.1.1) precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 20-Jan-1996 #sequence revision 25-Apr-1997 #text\_change 11-Jan-2002  
C:Accession: S68252; S72393; JC4391; S06440  
R:Jitrapakdee, S.; Booker, G.W.; Cassidy, A.I.; Wallace, J.C.  
Biochem. J. 316, 631-637, 1996  
A:Title: Cloning, sequencing and expression of rat liver pyruvate carboxylase.  
A:Reference number: S68252; MUID:96257760; PMID:8687410  
A:Accession: S68252  
A:Molecule type: mRNA  
A:Residues: 1-1178 <JIT1>  
A:Cross-references: EMBL:U36585; NID:g1040973; PIDN:AAC52668.1; PID:g1040974  
A:Accession: S72393  
A:Molecule type: protein  
A:Residues: 489-505 <JIT2>  
A:Experimental source: liver  
R:Lehn, D.A.; Moran, S.M.; MacDonald, M.J.  
Gene 165, 331-332, 1995  
A:Title: The sequence of the rat pyruvate carboxylase-encoding cDNA.  
A:Reference number: JC4391; MUID:96096548; PMID:8522203  
A:Accession: JC4391  
A:Molecule type: mRNA  
A:Residues: 1-221, 'P', 223-865, 'D', 867-976, 'G', 978-1178 <LEH>  
A:Cross-references: GB:U32314; NID:g929987; PIDN:AAA96256.1; PID:g929988  
A:Experimental source: liver  
R:Thampy, K.G.; Huang, W.Y.; Wakil, S.J.  
Arch. Biochem. Biophys. 266, 270-276, 1988  
A:Title: A rapid purification method for rat liver pyruvate carboxylase and amino acid  
A:Reference number: S06440; MUID:89024676; PMID:3178228  
A:Accession: S06440  
A:Molecule type: protein  
A:Residues: 'SG', 23-25, 'PL', 28-29, 'LL', 32-34, 'P', 1134, 'A', 1136-1137, 1139-1152, 'T', 1154-1155  
C:Comment: This enzyme is located in the mitochondrial matrix and catalyzes the conversion of pyruvate to oxaloacetate. It is a heterodimeric enzyme consisting of two subunits, alpha and beta. The alpha subunit is encoded by the JC4391 gene and the beta subunit is encoded by the S06440 gene. The enzyme is involved in gluconeogenesis and is a key enzyme in the regulation of glucose metabolism.  
C:Genetics:  
A:Gene: pc  
C:Superfamily: pyruvate carboxylase; biotin carboxylase homology: lipoyl/biotin-binding  
C:Keywords: biotin binding; gluconeogenesis; homotetramer; ligase; mitochondrion  
F:1-20/Domain: transit peptide (mitochondrion) #status predicted <TNP>  
F:21-1178/Product: pyruvate carboxylase #status predicted <MAT>  
F:339-494/Domain: biotin carboxylase homology <BCH>

F:1105-1178/Domain: lipoyl/biotin-binding homology <LPB>  
F:1144/Binding site: biotin (Lys) (covalent) #status experimental

Query Match 43.5%; Score 2548.5; DB 2; Length 1178;  
Best Local Similarity 46.3%; Pred. No. 3.5e-132;  
Matches 545; Conservative 188; Mismatches 409; Indels 35; Gaps 13;

QY 7 GGLLLKGIITLVSTHTSSTLPA-----FKKILVANRGEIATVAFRAALHTGAA 54  
Db 9 GGLRLGV-----RRSSTAPVSPNVRLEYKPIKKVWVANNRGEIATVFRACTELGR 62  
QY 55 TVAIYPREDRGSFHRFSASEAVRIGTEGSPVKAYLIDIDEIIGAARKVKADAIYPGFGLS 114  
Db 63 TVAVYSEQDTGMHRQKADAEVILGRGLAPQAYLHPDIKIVAKENGVDVAVHPGFGLS 122  
QY 115 ENAQLARECAENGITFIGTPEVLDLTGDKSRVTAAKKAGLPVL-AESTPSKNIDIVK 173  
Db 123 ERADEAQAQODAGVRFIGSPSEVVRKMGDKVEARATAIAAGVPPVPTGPTNSLSEAE 182  
QY 174 SAEGOTYPIFVKAVAGGGGRMFVSSPDRLRKLATEASREAEAFAGDSVVYVERAVINP 233  
Db 183 FNTYGFPIIFKAAVGGGGRVHVHSYELEENYTRAYSEALAAFGALFVEKEIEKP 242  
QY 234 OHIEVOILGDRTEGVVHLRYERDCLQRRHQKVVVEIAPQAHLDPDLRDRICADAKVFCRSI 293  
Db 243 RHIEVQILGDQYGNILHLVERDCSIQRRHQKVVVEIAPATHLDPQLSRSLTSDSVKLAKOV 302  
QY 294 GYOGAGTFEVLDEKGNHVFIEMNPRIQVHVTEVTEVDLVKAQMLAAGATLKLGL 353  
Db 303 GYENAGTFEVLVDKHKHYFIEVNSRLQVHTVTEITDVLVHAQIHYSEGSLPDLGL 362  
QY 354 TODKIKTHGAALQCRITTEDPNNGRPDGTITAYSPGAGVRLDGAQAL-GGETAHF 412  
Db 363 ROENIRINCAIQCRVTTEDPARSQPDITGRTEVFRSGEGMIRLONASAFQGANVLSPHY 422  
QY 413 DSNLVKMTCRGSDFFETAVARAQALAEFTVSGVATNIQGLRLLRREDDEFTSKRIATGFTG 472  
Db 423 DSNLVKVIANGKDHPTAATKMSRALAEFRVGVKTNIPFLQNVNNOQLAGIVDTQFID 482  
QY 473 DHPHLLQAPPADDEOGRILLDYLVNTVKNPHGVPRKDA-APIDKL-PNIKOLPLPRGSR 530  
Db 483 ENPELQLRPAQNRRAQKLLHLYGHVMVNGTPTPIPVKVSPPVDPVIVPVVPTGPPGPR 542  
QY 531 DRLKQLGPAAFARLREQDALAVTDTFRDAHQSLIATRVRSFALKPAAEAVALKTPPELL 590  
Db 543 DILLREGPEGFARAVNHQGLLLMDTTFRDAHQSLIATRVTHDLKXIAPYAHNPNFLF 602  
QY 591 SVEANGATYDVAMRFLFDPDWRDLDELREAMPNVNIQMLRGRNTVGTYPYPSVCRAF 650  
Db 603 SIENMGATFDVAMRFLYECPNRRLQELRELIPNIPFQMLLRGANAVGVTNYPDNVVKF 662  
QY 651 VKEAASSGVDFIERIDALNDVQMRPAIDAVLETNTAVAEVAMAYSGDLSDPNEKLYTLD 710  
Db 663 CEVAKENGMDVFRIFDSNLNLPNMLLGMEEAGSAG-GVVEAAISYTGVDADPSRTKYSLE 721  
QY 711 YVLKMAEELVKGHAHLAKDMAGLLRPAATVTKLVTALRREF-DLPVHVHTHTDTAGQLA 769  
Db 722 YVGLAEELVRAGTHLCLIKDMAGLLKPAACITMLVSSLRDRFPDPLHITHHTDTSSGVA 781  
QY 770 TVFAAQAQADAVDGAASAPLSGTTSPSLSAIVAFAHTRDTGSLSEAVSDLEPYWEAV 829  
Db 782 AMLACAQAGADVVDVDSMGSTSPSMGALVACTKGTPLDTEPLERVDSEYWECA 841  
QY 830 RGLYLPFESTGPTGTR--VYRHEIPGGQLSNLRAQATAGLADRELLIEDNYAANVEML 887  
Db 842 RGLYAAFDCTATMKSGNSDVSNEYENIPGGQYTNLHFQAHSMGLSGSKFKEVKAYVEANQML 901  
QY 888 GRPTKVTSPSKVGVGLALHLVAGVDPADFAADPOKYDIPDSVIAFLRGLGNPPGWPE 947  
Db 902 GDLIKVTSPSKVIGDLAQFVWQNGLSRAEAEAEELSPPRSVVEFLOQYIGIPHGFFPE 961  
QY 948 PLRTRAL-----EGRSEKGAPLTEVPEEEQAHLDA-DDSKERRNSLNRLLPKPTPEEL 1000



A:Cross-references: GB:299111; GB:AL009126; NID:G2633699; PIDN:CAB13359.1; PID:G2633857  
A:Experimental source: strain 168

C:Genetics:

A:Gene: PYCA

C:Superfamily: pyruvate carboxylase; biotin carboxylase homology; lipoyl/biotin-binding

C:Keywords: ligase

F:8-465/Domain: biotin carboxylase homology <BCH>

F:1073-1146/Domain: lipoyl/biotin-binding homology <LPB>

F:1112/Binding site: biotin (lys) (covalent) #status predicted

Query Match 43.1%; Score 2530; DB 2; Length 1148;

Best Local Similarity 46.0%; Pred. No. 3.6e-131;

Matches 533; Conservative 188; Mismatches 390; Indels 48; Gaps 13;

QY 28 AFKKILVANRGEIAVRAFAAALTAETGAATVAIYPRDRGSGFHRSEAFSEAVRIGTEGSPVKA 87

DB 5 SIQKLVANRGEIAIRFRACTELNRTVAVYSKEDSGSYHRYKADEAYLVGEGKPIDA 64

QY 88 YLDIDIEIIIGAAKKVADATPYGGLFSENAQLARECAENGITFPGTPEVLDLTGDKSRA 147

DB 65 YLDIEGIIIDIAKRNKVDATHPGYGLFSENIHFARRCEEIGVIFGPKSEHLDMEFGDKVA 124

QY 148 VTAACKAGLPVLAEST-PSKNIDDDIVKSAEGQTYPIFVKAVAGGGGRGMRVSSPDRLK 206

DB 125 REQAEKAGIPVTPGSDGPAETLEAVEFQFGQANGYPYIIKASLGSGGGRMRIVRSESEVKE 184

QY 207 LATEASREAEAAFGDGSVYVERAVINPOHIEVOIGLDRTEGVVHLYERDCSLQRRHOKYV 266

DB 185 AYERAKSEAKAFNDENVIEKLIENPKHIEVOIGDKGNVHLYERDCSVQRRHOKVI 244

QY 267 EIAPQHLDPLEDRICADAFKCRSIGYOGATVBELVDEKGNHVFIEEMNPRIQVETRV 326

DB 245 EVAPSVLSPELRDQICEAVALAKNVNINAGTVSEFLV-ANNEFYFIEVNPVQVETI 303

QY 327 TEEVTEVDLVKAQMLRAAGATL--KELGLTQDK-IKTHGAALOCRTITTEDPNNRFPDTC 383

DB 304 TEMITGVDIVQYQILVAQGHSLHKKVNIPEQKDIETIGYATQSRVTTEDPQNDPMDTG 363

QY 384 TITAYRSPGGAGVRLD-GAAQLGGEITAHFDSMLVMTCTGSGDFETAVARAQALAEFTV 442

DB 364 KIMAYRSGGGFVRLDTGNSFGAVITPYDLSLVKLTWALTFEQAQAAKKVNRNLQEFRI 423

QY 443 SGVATNIGELRALLREEDTSKRATGFTGDHPHLLQAPPADDEQGRILDYADVTVN-- 500

DB 424 RGKTNIPLENVAKHEKLTQGYDSFIDTTPLEFNPFPKQDRGFKMLTYIGNVTVNGF 483

QY 501 -----KPHGVRPKDVAAPIDKLPNIKDLPLPRGSRDLKQLGPAFAFARDLREQDALAV 553

DB 484 PGIGKKEKPAFDKPLGVKVDVQDP-----ARGTKQILDEKGAEGLANWVKEQKSVLL 536

QY 554 TDTTFDRAHQSLATRVRSFALKPAAEAVAKLTPELLSVEAMGATYDVAMRPLFEDPDWD 613

DB 537 TDTTFDRAHQSLATRIHRSHDLKIANPTAALWPELFSMEMGGATFDVAYRFLKEDPWK 596

QY 614 RLDELREAMPNNIOMLLGRNTGVTPYDSCVRAFEVKEAASSGVVDIFRFDALNDVSQ 673

DB 597 RLEDLRKEVPNTLFQMLRSSNANGVTNPDNVIKEFVKQSQSGDVFRIEDSLNNVKG 656

QY 674 MRPAIDAVLETNVAEAVAMAYSGDLSDPNEKLYTLDYILKMAEEIVKSAHILAIKDMA 733

DB 657 MTLAIDAVRDTG-KVAEAAICYGDIIDKRNKYDIADYTSMAKEALEAGATLGLKDMA 715

QY 734 GLLRPAAVTKLVTALRREEDLPVHVHTHTDAGGQATYFAAAQAGADAVDGASAPLSGTT 793

DB 716 GLLKPOAAVELYSALKETDIPVHLHTHTDSNGIYMYAKAVEGVDDIIDVAVSSNAGLT 775

QY 794 SQPSLSAIVAAFAHTRDRDTGLSLEAVSDLEPVEAVRGVLYLPESCTPGTGKRVYRHEIP 853

DB 776 SQPSASGCFHAMEGNDRRPMNVQVGLLSQYVESYKVIYSEFSGMKSPHTEIYHEMP 835

QY 854 GGOLSNLRAQATLGLADRFELLEDNAAVNEMLGRPTKVTPTSSKVVGDALHLVAGVD 913

DB 836 GGOYSNLQOQAAKGVGLGRWNEVYKEMRYRNDMFGDIVKVTPTSSKVVGDALYVQNNLT 895

QY 914 PADFAADPOKYDIPDSVIAFLRGELGNPPGWPPEPLRTRALEGRSECKAPLTVPEP--- 970

DB 896 EKDVYKESGLDPPDSVVELFKNGINQPHGFFPEKQKLKLGQE-----PIYVRPGELLE 951

QY 971 -----BQAHLDADDSDKERNLSNRLFLFPKPTPEEFLEHRRFRFGNTSALDDREFF 1018

DB 952 PVSFEALKQEFKEHNLISD----QDAVAYALYPKVFTDYVKTTSYGDISVLDTPTEFF 1007

QY 1019 YGLVEGRETLRLPLDVRTPLLVRLDAISEPDDKGMNRVANVANVGQIRPMRVDRSVEVST 1078

DB 1008 YGMLTGEETEVEVERGKT-LIVKLISIGEPQDPATRVVYFELANGQPREVVIKDESICSSV 1066

QY 1079 ATAEEKADNKGVAAPFAGVTVTVAE-GDEVKAGDAVAIIEMAKMEATITASVQDKIE 1137

DB 1067 OERLKADRTNPISHAASMPGTVIKVLAEAGTKVNGKGDHLMINAMKMETTVQAPFSSTIK 1126

QY 1138 RVVVPAAATKVEGGDLIVVV 1156

DB 1127 QVHVKNGEPIQTGDLLEI 1145

RESULT 6

JC2460

pyruvate carboxylase (EC 6.4.1.1) precursor - human

N:Alternate names: pyruvate:carbon dioxide ligase (ADP-forming)

C:Species: Homo sapiens (man)

C:Date: 10-Sep-1999 #sequence\_revision 18-Aug-2000 #text\_change 01-Feb-2002

C:Accession: G01933; JC2460; B27883; S01469

R:Walker, M.E.; Jitrapakdee, S.; Val, D.L.; Wallace, J.C.

submitted to the EMBL Data Library, July 1995

A:Reference number: H00708

A:Accession: G01933

A:Status: translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-1178 <WAL>

A:Cross-references: EMBL:U30891; NID:g1101028; PIDN:AAAB2937.1; PID:g1101029

R:MacKay, N.; Rigat, B.; Douglas, C.; Chen, H.S.; Robinson, B.H.

Biochem. Biophys. Res. Commun. 202, 1009-1014, 1994

A:Title: cDNA cloning of human kidney pyruvate carboxylase.

A:Reference number: JC2460; MUID:94324922; PMID:8048912

A:Accession: JC2460

A:Molecule type: mRNA

A:Residues: 1-224,'wp',227-351,'A',353-384,'PT',387-485,'DV',488-637,'R',639-728,'A',

A:Cross-references: GB:S72370; NID:G632807; PIDN:AAB31500.1; PID:G632808

R:Lambonvahi, A.M.; Quan, F.; Gravel, R.A.

Arch. Biochem. Biophys. 254, 631-636, 1987

A:Title: Sequence homology around the biotin-binding site of human propionyl-CoA carb

A:Reference number: A27883; MUID:87212051; PMID:3555348

A:Accession: B27883

A:Molecule type: mRNA

A:Residues: 1083-1178 <LAN>

A:Cross-references: GB:M26122; NID:g189657; PIDN:AAA36423.1; PID:g387003

R:Freytag, S.O.; Collier, K.J.

J. Biol. Chem. 259, 12831-12837, 1984

A:Title: Molecular cloning of a cDNA for human pyruvate carboxylase. Structural rel

A:Reference number: S01469; MUID:85030380; PMID:6548474

A:Accession: S01469

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1135-1178 <FRE>

A:Cross-references: EMBL:K02282; NID:g189655; PIDN:AAA60033.1; PID:g189656

C:Genetics:

A:Gene: GDB:PC

A:Cross-references: GDB:119472; OMIM:266150

A:Map position: 11q11-11q13.1

C:Superfamily: pyruvate carboxylase; biotin carboxylase homology; lipoyl/biotin-binc

C:Keywords: biotin binding; gluconeogenesis; ligase; mitochondrion

F:1-20/Domain: transit peptide (mitochondrion) #status predicted <TNP>

F:21-1178/Product: pyruvate carboxylase #status predicted <NAT>

F:39-494/Domain: biotin carboxylase homology <BCH>

F:1105-1178/Domain: lipoyl/biotin-binding homology <LPB>

F:1144/Binding site: biotin (lys) (covalent) #status predicted

```
Query Match 43.1%; Score 2528.5; DB 1; Length 1178;
Best Local Similarity 45.9%; Pred. No. 4.5e-131;
Matches 540; Conservative 187; Mismatches 415; Indels 35; Gaps 13;

QY 7 GGLLKGITLIVSTHTSTLPA-----FKILVANRGEIAVRAFRALLETGAA 54
DB 9 GGLRLGI-----RRTSTAPAAAPNVRRLLEYKPIKKVMANRGEITAIRFRACTELGIR 62
QY 55 TVAIYPREDGSRFIRSPASEAVRIGTSGSVKAYLDIDELIGAAKVKADAIYPGVGFLS 114
DB 63 TVAIYSEDQTCQMRQADEAYLIGRLAPVQALHPDILIKVAKENNVDAVHPGFGFLS 122
QY 115 ENAOLARECANGTITGPPEVDLDLFGDKSRAVTAAKKAGLPVL-AESTPSKNIDIVK 173
DB 123 BRADFAQACQAGVRFTIGPSEVVRKMGDKVEARAIAAGVPVPGTDPATIPYSLHEAHE 182
QY 174 SAEGOTPIFYKAVAGGGGMRVSSPOELRKLATPASEAREAAAFDGGSVYVERAVINP 233
DB 183 FSNTYGFPIIFKAAAYGGGGRMVVHSYEELNNTYRAYSEALAAFGNGALFVEKEFERP 242
QY 234 OHIEVOILGDRTEGVHLYERDCSLQRHQRHVVEIAPQAHLDPRLDRICADAVKFCRSI 293
DB 243 RHIEVOILGDOYGNLHLYERDCSIQRHQRHVVEIAPQAHLDPOLRLTSDSVKLAKOV 302
QY 294 GYQAGATVERVLVDKGNHVFTEMPRIQVHTVTEVTEVDLVKAQMLAAGATLKLGL 353
DB 303 GYENAGTVEFLVDRGHKHYFEVNSRLQVHTVTEITDVLVHAQTHVSEGRSLPDLGL 362
QY 354 TODKTKHGAALOCRTITTEDPNNGFRPDTGITATAYRSPGGAGVRLDGAAL-GGEITAFH 412
DB 363 RQENIRINGCAIQCVTTEDPARSPQDTGRIEVRSGEGMIRLDNASAFQAGAVISPHY 422
QY 413 DMLVKWTCRSDPETAVARAORALAEFTVSGVATNIGFLRALLREDFTSKRITATGFIG 472
DB 423 DSLLVKTAHGKDHPATKMSRALAEFRVGVNTKNTAFIQLNNOQFLAGIVDTGIFD 482
QY 473 DHPHLLQAPPADDEGRILDLADVTYVKNKPHGVRP-KDVAAPDK-LPNKIDPLPRGSR 530
DB 483 ENPELFQURPAONRAOKLHLVGHVWNGPTPIPVKASPSPTDPVVPVPIGPPAGFR 542
QY 531 DRLKOLGPAAPARDLEODALAVDTTFRDAHQSLLATRVRSFALKPAAEAVAKLTPELL 590
DB 543 DILLREGPEGFARAVRNHPGLLMDTTFRDAHQSLLATRVRTDHLKKTAPVVAHNFSKLF 602
QY 591 SVEAWGATYDVAMRFLFEDPDWDLDELREAMPNVNQMILLGRNTYGYTYPDVSVCRAF 650
DB 603 SMENWGGATFDVAMRFLYECPRRLQELRELIPNIPEQMLLRGANAVGYTYNPDNVYFKF 662
QY 651 VKEAASGVDIFRFDALNDYSQMRPAIDAVLETNTAAVEAMAYSGDLSDPNEKLYTLD 710
DB 663 CEVAKENGMDVRFVDSNLNPLNLGMEAGSAG-GVVEAAISYTGADVADPSRTKYSLQ 721
QY 711 YLKWAEETVKSAGHILAIKDMAGLLRPAAYTKLVTLRREF-DLPVHVITHDTAGGOLA 769
DB 722 YMGLAEBELVRAGTHILICIKDMAGLLKTACTMLVSSLRDRFPDPLHIHTHDTSGAGVA 781
QY 770 TYFAAQAQADAVDASAPLSTGTSQSLSAIVAFAAFAHTRDTGLSEAVSDLEPYKEAV 829
DB 782 AMLCAQAQADVVDVADSMGSMGTSQSPMGALVACTRGTPLDTEVPMEVDFYSEYEGA 841
QY 830 RGLYLPFESGTPGPTGR--VYRHEIPGQLSNLRAQATALGLADRFELIEDNYAAVNEML 887
DB 842 RGLYAAFDCTATMKSNGSDVYENETPGQYTNLHFQAHSMGLGSKFKFEVKAYVEANQML 901
QY 888 GRPTKVTPTSSKVVGDALHLVAGVDPAADFAADPOKYDIPDSVIAFLRUGELGNPPGWPE 947
DB 902 GDILKVTPTSSKIVGLDAQFMVQNGLSRAEAQAQAEELSFRPSVYVEFLQGVIGVPHGFFE 961
QY 948 PLURTRAL-----EGRSEKCAPLVEVPEEQAHLDA--DDSKERNNSLNRLIFPKPTTEFL 1000
DB 962 PPRSKVLKDLPRVECRPGASLPDLQALEKEKELVDRHCEEVTPTDVLSAAMYPDVFAHFK 1021

QY 1001 EHRRFGNTSALDDREFFYGLVEGRETLIRLPDVRTPLLRDLAISEPDDKGMNRNVAVY 1060
DB 1022 DFTATFGPLDLSNTRFLTQGPRTAEFEVELERGKT-LHIKALAVSDLNRAQOROVFFEL 1080
QY 1061 NGQIRPMRVRDSVESVTATAEKADSSNKGHVAAPFAG-VVTVTVAEGDEVKAGDAVAII 1119
DB 1081 NGOLRSILVKDQAMKEMHFPKALKDKVKGIGAPMPGKVIDIKVVAGAKVAKGQPLCVL 1140
QY 1120 EAMKWEATITASVDCGKIERYVVPVPAATKVEGGDLIVVV 1156
DB 1141 SAKMKTVVTSPMEGTIVRKVHVTKDWTLBSGDDLLLEI 1177

RESULT 7
AH1208
pyruvate carboxylase homolog pycA [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AH1208
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AH1208
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1146 <GLA>
A:Cross-references: GB:NC_003210; PID:NCAC9150.1; PID:g16410474; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
C:Gene: pycA
C:Superfamily: pyruvate carboxylase; biotin carboxylase homolog; lipoyl/biotin-bindi

Query Match 42.5%; Score 2494; DB 2; Length 1146;
Best Local Similarity 45.7%; Pred. No. 3.4e-128;
Matches 525; Conservative 186; Mismatches 408; Indels 30; Gaps 11;

QY 30 KTLVANRGEIAVRAFRALLETGATVAIYPREDGSRFIRSPASEAVRIGTSGSPKAYL 89
DB 5 KVLVANRGEIAIRVMRACTELKIKTVAIYSEDQTSFHRYSDEAYLVGAGKGPIDAIL 64
QY 90 DIDEIIGAARKVKADAIYPGYFLSENAQLARECAENGITFTGPTPEVLDLTDGDKSRVAT 149
DB 65 DTENITIEAKESGADAIHPGYGFLSENIEFARRCEQEGIFVGPXSKHJDMFGDKTAK 124
QY 150 AAKKAGLPVLAEST-PSKNIDDIVKSAEGOTYPIFYKAVAGGGGMRVSSPDELRLKLA 208
DB 125 QALLADIPVPGSGVPVAGTKEVEEGERKNGYPLMTKASLGGGGGMRVSEKHKESF 184
QY 209 TEASREAAFGDGSVYVERAVINPOHIEVQILGDRTEGVHLYERDCSLQRHQRHVVEI 268
DB 185 ERASSEAKAAGFNDEYVEKCVNPKHIEVQILGDRTHGNIVHLFERDCSIQRHQRHVVEI 244
QY 269 APAQLHDPRLDRICADAVKFCISGYCAGATVEFLVDEKGNHVFTEMPRIQVHTVTE 328
DB 245 APCNAITSELNRICDAAVKLMKNVDYINAGTVEFLV-EGDDFYFLEVAPRVQVHTITE 303
QY 329 EVTEVDLVKRAOMLAAGATLKLGLT---QDKTKHGAALOCRTITTEDPNNGFRPDTGHI 385
DB 304 MITGIDIVOSQLFIADGYALHDOLVAIPKQEDIHGSAIQSKRITTEDPLNPNMPTDGRV 363
QY 386 TAYRSPGGAGVRLD-GAAQLGGSEITAHFDSMLVKMTCRGSDFETAVARAORALAEFTVSG 444
DB 364 DTYRSTGGGVRDLADAGNGFGVTVPYDLSLVKLTGWTGTMTEFQATRKMRNLIEFRIG 423
QY 445 VATNIGFLRALLREEDFTSKRIATGFIGDHPHLLQAPPADDEQGRILDLADVTYVKNKPH 504
DB 424 VKTNIPLLNVVRHPDFAFASNGYNTSFIDTTPELFKPPIHNDRTGKTGLRYIGNVTAVGFFG 483
```









```
Db 617 TEDPWERLSLIREGAPNLLQLMRLGANGVKNYPDMNVKVFYRQAAAGVDLRFVDC 676
Qy 668 LNDVSOQMPAIDAVLETNATAVAEMAYSGDLSDPNEKLYTLDYLRKMAEIVKSGAHIL 727
Db 677 LNWVENMVSMDATAEER-KICEATICYTGDLNLSARPKYDLKYTYTNLAWELEKAGAHII 735
Qy 728 AIKDMAGLLRPAAVTKLVTLALRRERFDLPVHVHTHTDAGGQLATYFAAAQAGADAVGASA 787
Db 736 AVKDMAGLLKPAKAAKLVKALREATGLPIHFHTHTDTSISAAATVLAADAVDAVDAAMD 795
Qy 788 PLSGTTSPSLSAIWAFAHRTDGLSLEAVSDLEPWEAVRGLYLPFESGTPGPTGRV 847
Db 796 AFSGNTSQCLGSIVEALSGSERDGLDTEWIRLSFWEAVRNOYAAAFESDLKGPASEV 855
Qy 848 YRHETPGQLNLRAQATLGLADREFELIDNYAAVNEMLGRPKTKVPSKVVGDIALHL 907
Db 856 YLHEMPGGOFTNLKEQARSLSLESHWEVAQYADANRMEGDIKVPVPSKVVGDMLMM 915
Qy 908 VGAGVDPADFAADPKYDIPDSVTAFLRGLNPPGPGWPEPLRTRALEGRSEKAPLITEV 967
Db 916 VSODLTAVDENPKEVSFSDVSMKGLDQSGPGWPEALQKAL----KGEKPYTVR 971
Qy 968 PEE--EQAHLDAADSKERNLSNR-----LLFPKTEEFLEHRRFRGNTSALDDR 1015
Db 972 PCSLLEDADLDA-ERKVIETKLERKDDFEFASVLMYPKVFTDFALTAETVGPVSVLPTH 1030
Qy 1016 EFFYGLVGEQRETLIRLPDVRTPLVRLDAISEPDDKGMNRYVANVGOIRPMVRDRS-V 1074
Db 1031 AIFYGMEDEGEELFADIERGKTLVIN-QASSGIDDKGMVTVFEEINGOPRRIKVPDRAHG 1089
Qy 1075 ESVTATAEKADSSNKGHVAAPFAGVVT-VTVAGDEVKAGDAVAIIEAMKMEATITASVD 1133
Db 1090 ASGSVRRKAEPCGNASHICAPMPGVISRVFINOGQEVKAGDVLLSIEAMKMETALHAERD 1149
Qy 1134 GKIERVVVPAATKVEGGDLIV 1154
Db 1150 GKIAEVLVKPGDQIDAKDLII 1170

RESULT 12
AE2911
pyruvate carboxylase [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AE2911
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H.;
  erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,
  Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
  ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AE2911
A:Status: preliminary.
A:Molecule type: DNA
A:Residues: 1-1174 <KUR>
A:Cross-references: GB:AF008688; PIDN:AAL43707.1; PID:g17741236; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: pyca
A:Map position: circular chromosome
C:Superfamily: pyruvate carboxylase; biotin carboxylase homology; lipoyl/biotin-binding

Query Match 42.2%; Score 2474.5; DB 2; Length 1174;
Best Local Similarity 46.5%; Pred. No. 4.2e-128;
Matches 540; Conservative 174; Mismatches 406; Indels 41; Gaps 19;

Qy 25 TLPAFKKILVANRGEIARAFRALETKAATVATVPREDRGSFHRFSASEAVRIG----- 79
Db 20 TVLKISKILVANRSEIARVFRANELGIKTVAIWAEDKLSLHRFRKADESYQVGRPHL 79
Qy 80 -TEGSPVKAYLIDDEIIGAANKVKADAIYPGYFLSENAQLARCAENGLTFIGTPEVL 138
```

```
Db 80 AKDMGIESYLSIEEIVIRVAKLSGADAIHPGYGLLSSEFVEACNKAGITIGPTPDTM 139
Qy 139 DLTGDKSRVATAAKKAGLPVLAESTP-SKNIDDIVKSAEQTYPIFVKAVAGGGGRMRF 197
Db 140 RQLGNKVAARNLAISVDVPVVPATNPLPDDIAEVEERMAEIGYPMVLMKASGGGGRMA 199
Qy 198 VSSPDELRLKATEASREAAAFGDSVYVERAVINPOHIEVOILGDRTEVHVHLEDCS 257
Db 200 IRKKEDLAREVTEAKREAAFGKDEVLKLVERRARHVESOILGDTGNVNVHLFERDCS 259
Qy 258 LORRHOKVVVEIAPOAHLDELDRICADAVKFCRSIGYOGAGTVPELVD-ERGNHVFICM 316
Db 260 IQRNKQVVERAPAPYLSEAQROELAAYSLKIAAATNYICAGTVEYLMADATGKFYIEV 319
Qy 317 NPRIQVHTVTEVTEVDLVKAQMLAAAGATL--KELGL-TQDKIKTHGAALOCRTITTD 373
Db 320 NPRIQVHTVTEVVTGIDIVKAQIHLEGAAGTAGESVPKQEDIRLNGHALOCRTITTD 379
Qy 374 PNNGRPDGTITAYRSPGAGVRLD-GAAQGLGETTAHFDSMLVKMTCRGSDFFETAVAR 432
Db 380 PEHNFIPDYGRITAYRSASGFIRLDGGYSYTGAVITRYIDPLLVKVTAWAPEPDEAIR 439
Qy 433 AQALAEFTVSGVATNIGELRALLREEDFTSKRIATGFIQDHPHLLQAPPADDEQGRILD 492
Db 440 MDRLALREFRIRGVATNLTFLEAIIGHDSFRNNTYTRFDISTPELPAQVKRODRATKLLT 499
Qy 493 YLADVTVN--KPHG-VRPKDVAA-PIDKLPNIKOLPLPRGSRDRQLKQLGPAFAFARDLRE 547
Db 500 YLADVTVNGHPETKGRAPKADKAAPKPI--VPYI-DAPTPDGTQKLLDKLPGKGFADMMN 556
Qy 548 ODALAYTDTTFRDAHOSLLATRVRSFALKPAEAAVAKLTPELLSVBAMGATYDVAMREL 607
Db 557 EKRVLVDTTMRDGHOSLLATRVTHDIARVASVYKALPQLLSLCEWGATFDVSMREL 616
Qy 608 FEDPWRLDELREAMPNVIQMLLRNTGVTPYPSVCRAFPAKAAASGVDIFRIDA 667
Db 617 TEDPWERLSLIREGAPNLLQLMRLGANGVKNYPDMNVKVFYRQAAAGVDLRFVDC 676
Qy 668 LNDVSOQMPAIDAVLETNATAVAEMAYSGDLSDPNEKLYTLDYLRKMAEIVKSGAHIL 727
Db 677 LNWVENMVSMDATAEEN-KICEATICYTGDLNLSARPKYDLKYTYTNLAWELEKAGAHII 735
Qy 728 AIKDMAGLLRPAAVTKLVTLALRRERFDLPVHVHTHTDAGGQLATYFAAAQAGADAVGASA 787
Db 736 AVKDMAGLLKPAKAAKLVKALREATGLPIHFHTHTDTSISAAATVLAADAVDAVDAAMD 795
Qy 788 PLSGTTSPSLSAIWAFAHRTDGLSLEAVSDLEPWEAVRGLYLPFESGTPGPTGRV 847
Db 796 AFSGNTSQCLGSIVEALSGSERDGLDTEWIRLSFWEAVRNOYAAAFESDLKGPASEV 855
Qy 848 YRHEIPGQLSNLRAQATLGLADREFELIDNYAAVNEMLGRPKTKVTPSKVVGDIALHL 907
Db 856 YLHEMPGGOFTNLKEQARSLSLESHWEVAQYADANRMEGDIKVPVPSKVVGDMLMM 915
Qy 908 VGAGVDPADFAADPKYDIPDSVTAFLRGLNPPGPGWPEPLRTRALEGRSEKAPLITEV 967
Db 916 VSODLTAVDENPKEVSFSDVSMKGLDQSGPGWPEALQKAL----KGEKPYTVR 971
Qy 968 PEE--EQAHLDAADSKERNLSNR-----LLFPKTEEFLEHRRFRGNTSALDDR 1015
Db 972 PCSLLEDADLDA-ERKVIETKLERKDDFEFASVLMYPKVFTDFALTAETVGPVSVLPTH 1030
Qy 1016 EFFYGLVGEQRETLIRLPDVRTPLVRLDAISEPDDKGMNRYVANVGOIRPMVRDRS-V 1074
Db 1031 AIFYGMEDEGEELFADIERGKTLVIN-QASSGIDDKGMVTVFEEINGOPRRIKVPDRAHG 1089
Qy 1075 ESVTATAEKADSSNKGHVAAPFAGVVT-VTVAGDEVKAGDAVAIIEAMKMEATITASVD 1133
Db 1090 ASGSVRRKAEPCGNASHICAPMPGVISRVFINOGQEVKAGDVLLSIEAMKMETALHAERD 1149
Qy 1134 GKIERVVVPAATKVEGGDLIV 1154
Db 1150 GKIAEVLVKPGDQIDAKDLII 1170
```



QY 149 TAAGKAGLPVLAEST-PSKNIDDIIVKSAEGQTYPIFKAVAGGGGGRGMRVSPDELRLK 207  
Db 152 QAAIEAGVQVPGPGPTTTTDAEAVEFAKQYGTPIILKAAYGGGGRIRRVLDLEEVEEA 211  
QY 208 ATEASREAAEAFGDSGVYVERAVINPOHIEVQIILGDRGTGEVHLVYERDCSLQRHQRKVE 267  
Db 212 FRRSYSEAAAFGDSGLFVEFRPRHIEVQLLGDHGNVHLVYERDCSVQRHQRKVE 271  
QY 268 IAPAQHLDPDLDRICADAVKFCRSIGVQAGTVEFLVDEKGNHVFTEMPPRIQVEHTVT 327  
Db 272 IAPALPEGVREKLADALRLARHVQYAGTVEFLVDQKGNVYFIEVNARLQVEHTVT 331  
QY 328 EEVTEVDVLKQMRLAAGATKELGLTQDKTKTHGAALQCRITTEDPNNNGFRPDGTGITA 387  
Db 332 BEITGVLDVQAOIRIABEGLSDDLKLSQETIQTGSAIQCRVTTEDPAKGFQDPSGRIEV 391  
QY 388 YRSPGGAGVRLDGAALGGE-ITAFDPSMLVKYTCRSDSETAVARAQALAEFTVSGVA 446  
Db 392 FRSEGMGIRLDSASAFAGSVISPHYDSLAVKYIATARNHPNAAKMRALKFRIRGVK 451  
QY 447 TNIGFLRALLREDEFTSKRIATGFIGDHPHLLQAPPADEQGRILLYLADVTYVKNP- 502  
Db 452 TNIPFLLNVLRQPSFLQASVDYTFIDEHPOLFOPKPSONRAOKLLNLYLGEVKNPPTPL 511  
QY 503 -HGVRPKDVAAPIDKLPNIKDLPLPRGSRDLKQLGPAFAFARDLREQDALAVDTTFRDA 561  
Db 512 ATDLKPAVVSPIPIYIP--AGAKPTGLRDVLVQVGTPEFAKEVRSRPGCMITDTTFRDA 569  
QY 562 HQSLATRVRSFALKPAEAAVAKLTPELLSVFAMGGATYDVAMRFLPEDPDRDLRELA 621  
Db 570 HQSLATRVRYDMAAISPFAQFNGFLFSLNMGATFDVSMRFLHECPWERLQTLRLK 629  
QY 622 MPNNIQMLLRGNTVGYTPYPSVCFRAVKEAASSGVDFRIFDALNDVYSQMRPALDAV 681  
Db 630 IENIPFOCLLRGANAGYSNYPDNVYKFCELAVKMGDMVFEVFDLSNLYPLNLLVGM 689  
QY 682 LETNTAVAEVAMVSGDLSDPNEKLYTLDYVYKMAEIVKSGAHILAIKDMAGLLRPAV 741  
Db 690 GKAG-GVVEAAIAVTDGVTDSKREKYLKYYLNLADQLVRAQAHILSIKDMAGVLRPEA 748  
QY 742 TKLVTALRRP-DLPVAVHHTDTAGGOLATYFAAAQAGADAVDAGASAPLSTGTSQPSLSA 800  
Db 749 KLLICALRDKFPDPIHVHTHTDSAGVAAMLECAKAGADVDAVDAVDSMGMTSQPSMGA 808  
QY 801 IVAFAHTRRTGLSLEAVSDLEPYEAVRGLYLPFESGTPGPTGR--VYRHEIPGQOLS 858  
Db 809 IVASLOGTKHOTGLSDDISKYSAYWESTKOLYAPFECATMKSGNADVYKHEIPGQYT 868  
QY 859 NLRAQATALGLADRFELIEDNYAAVNMELGRPTKVTPTSPKSVGDLALHLVAG-----V 912  
Db 869 NLQFAFSLGLGQPFQDEVKRYREANLVGLDIIKVTPESSKIVGDLAQFMVQNNLTRTLV 928  
QY 913 DPADFAADPQKIDIPDSVIAFLRGELGNPPGWPPEPLRTRALEGRS--EGKAPLVEYPEE 970  
Db 929 DRAD-----DLSPFKSVVDMQGNVGPYPYGFPEPLRTKVLGRKPKVDGR-----PGE 976  
QY 971 EQAHLDAADSK-----ERNRSLNLLFPKPTTEFLHRHRRFGNTSALDDREFFY 1019  
Db 977 NAKPVLDLAVKVELEKHGRLTSEEDVMSYSMPTPTVEDEFETROQYGPVDKLPRLFLT 1036  
QY 1020 GLVEGRETILRLPDVRYPLVRLDAISEPDDKGMNVVAVNVNNGQIRPMRYRDRSVESVTA 1079  
Db 1037 GLEIAEEVDVEIESKT-LAIQLLAEGKLNKRGEREVFFDLNGQMRISFVVDKEASKEIV 1095  
QY 1080 TAEKADSSNKGHVAAPFAG-VVTVTVAGDEVKAGDAVAIIIAEMKMEATTITASVQCKIER 1138  
Db 1096 TRPALPQVRGHICAPMPGDVLEKIKEGDKVYKKQPLFVLISAKMDEMVIDSPIAGTVKA 1155  
QY 1139 VVVAATKVEGGDLIVVV 1156  
Db 1156 IHAPQGTNCSAGDLVVEV 1173

RESULT 15  
QYBYB  
pyruvate carboxylase (EC 6.4.1.1) 1 [validated] - yeast (Saccharomyces cerevisiae)  
NAlternate names: protein G3428; protein YGL062w; pyruvic carboxylase  
C:Species: Saccharomyces cerevisiae  
C:Date: 31-Dec-1991 #sequence\_revision 19-Jul-1996 #text\_change 01-Feb-2002  
C:Accession: S64066; A29233; S05722  
R:Feuermann, M.; Potier, S.; Souciet, J.L.  
submitted to the Protein Sequence Database, May 1996  
A:Reference number: S64044  
A:Accession: S64066  
A:Molecule type: DNA  
A:Residues: 1-1178 <FEU>  
A:Cross-references: EMBL:272584; NID:gl322565; PIDN:CAA96765.1; PID:gl322566; GSPDB:G  
A:Experimental source: strain S288C  
J.Lim, F.; Morris, C.P.; Occhiodoro, F.; Wallace, J.C.  
J. Biol. Chem. 263, 11493-11497, 1988  
A:Title: Sequence and domain structure of yeast pyruvate carboxylase.  
A:Reference number: A92662; MUID:88298805; PMID:3042770  
A:Accession: A29233  
A:Molecule type: DNA  
A:Residues: 1-461, 'G', 463-492, 'D', 494-594, 'A', 596-618, 'O', 620-663, 'S', 665-771, 'R', 773  
A:Cross-references: EMBL:J03889; NID:gl72101; PIDN:AAA34843.1; PID:gl72102  
R.Morris, C.P.; Lim, F.; Wallace, J.C.  
Biochem. Biophys. Res. Commun. 145, 390-396, 1987  
A:Title: Yeast pyruvate carboxylase: gene isolation.  
A:Reference number: S05760; MUID:87241529; PMID:3036126  
A:Accession: S05760  
A:Molecule type: DNA  
A:Residues: 1003-1178 <MOR>  
A:Cross-references: EMBL:J03889  
A:Accession: A29722  
A:Molecule type: protein  
A:Residues: 1124-1149 <MOR2>  
C:Genetics:  
A:Gene: SGD:PYC1; PYV: MIPS:YGL062w  
A:Cross-references: SGD:S0003030; MIPS:YGL062w  
A:Map position: 7L  
C:Superfamily: pyruvate carboxylase; biotin carboxylase homology; lipoyl/biotin-biotin-  
C:Keywords: biotin binding; gluconeogenesis; homotetramer; ligase; zinc  
F:21,478/Domain: biotin carboxylase homology <BC>  
F:157-33/Domain: ATP/bicarbonate binding #status predicted <ATB1>  
F:353-468/Domain: ATP/bicarbonate binding #status predicted <ATB2>  
F:569-908/Domain: pyruvate binding #status predicted <PYR>  
F:1096-1169/Domain: lipoyl/biotin-binding homology <LPB>  
F:1135/Binding site: biotin (Lys) (covalent) #status experimental

Query Match 41.9%; Score 2457; DB 1; Length 1178;  
Best Local Similarity 46.6%; Pred. No. 3.8e-127;  
Matches 540; Conservative 175; Mismatches 400; Indels 44; Gaps 16;

QY 31 KILVANRGEIIVRAFAALTAETGAATVAIYFREDRGFSFHRSEAVRIGTEG--SPVKAY 88  
Db 21 KILVANRGEIPIRIPTAHLSNQTVAIYSHEDRLSTHKQKADAEAYVIGEVQYTPVGA 80  
QY 89 LDIDEIIGAANKVADAIYPGYGLFSENALARECAENGITFTGPTPEVLDLTGDKSRV 148  
Db 81 LAIDEIISIAQKHQVDFIHPGYGLFSENSEFADKVKVAGITWIGPPAEVIDSVGDKVSAR 140  
QY 149 TAAGKAGLPVLAEST-PSKNIDDIIVKSAEGQTYPIFKAVAGGGGGRGMRVSPDELRLK 207  
Db 141 NLAARAVTVPTGCTPCPIETVEALDFVNEYGYPIIKAAGGGGGRGMRVREGDDVADA 200  
QY 208 ATEASREAAEAFGDSGVYVERAVINPOHIEVQIILGDRGTGEVHLVYERDCSLQRHQRKVE 267  
Db 201 FORATSEATFAGNGCFVREFLDKPHIEVOLLADNHNHGVHLLFEDRCSVQRHQRKVE 260  
QY 268 IAPAQHLDPDLDRICADAVKFCRSIGVQAGTVEFLVDEKGNHVFTEMPPRIQVEHTVT 327  
Db 261 VAPAKTLPREVRDALILTDAVKLAKKEGYRNAGTAEBFLVDNQNHRHFIETNPRQVEHTIT 320  
QY 328 EEVTEVDVLKQMRLAAGATKELGLTQDKTKTHGAALQCRITTEDPNNNGFRPDGTGITA 387

Db 321 EEITGIDIVAAQIAAGASLPQLGLFODKITTGRFAIQCRITTEDPAKNFQPDGTGRIEV 380  
QY 388 YRSPGGAGVRLDGA-AOLGGEITAHFDSMLVKWTCRGSDFFETAVARAOPALAEFTVSGVA 446  
Db 381 YRSAGNGVRLDGGNAYAGTIIISPHYDSMLVWCSCSGSYEIVRRMIRALIEFRIRGVK 440  
QY 447 TNIGFLRALLREEDFTSKRIATGFIQDHPHLLQAPPADDEQGRILDYLADEVN---KP 502  
Db 441 TNIPELLTLLTNPVFEIGTYWTFIDDTPLQFQMYSSONRAQKLLHYLADVAVNGSSIKG 500  
QY 503 HGVRPKDVAAPIDKLPNTKDL-----PLPRGSRDLKOLGPAAPARDLREODALA 552  
Db 501 QIGLPEKLSNP--SYPHLHDAAGNVINVTKSAPPSGWQVLLKEGPAEFARQVROFNGLT 558  
QY 553 VTDTTFRDAHQSLLATRVRSFALKPAAEAVAKLTPELLSVEAWGGATYDVAMRFLFEDPW 612  
Db 559 LMDTTRDAHQSLLATRVTRTHDLATIAPTTAHALACRFALECGGGATFDVAMRFLHEDPW 618  
QY 613 DRLEDELREAMPNVNIOMLLGRNTYGYTPYDSCVCAFVKEAASSGVDFIRIFDALNDVS 672  
Db 619 ERLRLRSILVPNIPEQMLLRGANGVAYSSLPDNAIDHFVKQAKDNGVDIFRVFDALNDLE 678  
QY 673 QMRPAIDAVLENTAVAEVAMAYSGLSDPNNEKLYTLOYLLKMAEEIVKSGAHILAKDM 732  
Db 679 QLVGVGDVAVKKAG-GVVEATVCFSGMLQPGKK-YNLDYILEIAEKIVQMGTIILGIKDM 736  
QY 733 AGLLRPAAVTKLVATLAREEF-DLPVHVHTHTAGQLATYFAAAQAGADAVDGAAPLSG 791  
Db 737 AGTMPAPAAKLLIGSLRAKYPDLPPIHVHTHDSAGTAVASMTACALAGADVVDVAINMSG 796  
QY 792 TTSQPSLSAIVAAFAHTRDRTGLSLEAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHE 851  
Db 797 LTSQPSINALLASL-EGNIDTGINVHVRELDAYWAEMRELLYSCEFADLKGPDPEVYQHE 855  
QY 852 IPGGQLSNLRAQATLGLADRELELDNYAAVNEMLGRPTKVTSPSKVVGDLALHLVGAG 911  
Db 856 IPGGQLTNLLFOAQLGLGEQWAEKRAYREANYLLGDIVKVTPTSKVVGDLAQFMYSNK 915  
QY 912 VDPADFAADPOKYDIPDSVIAFLRGLGNPPGPGWPEPLTRALEGRSEKAPLTVPEPEE 971  
Db 916 LTSDDVRLANSLDFPDSVWDFEGLIGQPYGCFPEPFRSDVLRNK---RRKLTCPGLE 972  
QY 972 QAHLDAADDSKERNLSNRL-----LFPKPTEEFLEHRRFRFGNTSALDDREFFY 1019  
Db 973 LEPPFOLE--KIREDLQNRFGDVEDCDVASYNMYPRVYEDFKMRETYGDLVSLPTRSFLS 1030  
QY 1020 GLVEGRETLIRLPDVRTPLLVRLDAISEPDDK-GMRNVVANVNGOIRPMRVDRDSVESVT 1078  
Db 1031 PLETEEEIEVWTEQGT-LIILQAVGDLNKKTGEREVYFDLNGEMRKIRVADRSQKVET 1089  
QY 1079 ATAERKADSNKGHVAAPFAGV-VTVTVAGDEVKAGDAVAIIEAMKMEATITASVDGKIE 1137  
Db 1090 VTKSRADMDPLHIGAPMAGVIVEVVKHKSGLIKKQPVAVLSAMKMEMIISPSDGGQVK 1149  
QY 1138 RVVPAATKVEGGDLIVVV 1156  
Db 1150 EVFVSDGENVDSDLLVLL 1168

Search completed: March 26, 2003, 19:03:38  
Job time : 59 secs





GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 18:55:00 ; Search time 16 Seconds

(without alignments)  
2999.257 Million cell updates/sec

Title: US-09-974-973-2

Perfect score: 5865

Sequence: 1 MTATLGLLKGITLVT.....RVVPAAKTVGGDLIVVVS 1157

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match %	Query Length	ID	Description
1	2550.5	43.5	1178	1 PYC_MOUSE	Q05920 mus musculus
2	2534.5	43.2	1178	1 PYC_RAT	P52873 rattus norv
3	2531.5	43.2	1178	1 PYC_HUMAN	P11498 homo sapien
4	2457	41.9	1178	1 PYC1_YEAST	P11154 saccharomyc
5	2436	41.5	1180	1 PYC2_YEAST	P32327 saccharomyc
6	2393	40.8	1189	1 PYC_PICPA	P78992 pichia past
7	1079	18.4	501	1 PYCA_METJA	Q58626 methanococ
8	1032.5	17.6	447	1 ACCC_ANASP	Q06862 anabaena sp
9	1020	17.4	506	1 PYCA_ARCFU	O30019 archaeglob
10	996	17.0	491	1 PYCA_METTH	O27939 methanobact
11	950	16.2	448	1 ACCC_HAEIN	P43873 haemophilus
12	940	16.0	449	1 ACCC_ECOLI	P24182 escherichia
13	938	16.0	449	1 ACCC_ECO57	Q8x366 escherichia
14	921	15.7	449	1 ACCC_PSEAE	P37798 pseudomonas
15	897	15.3	725	1 MCCA_HUMAN	Q96rq3 homo sapien
16	894	15.2	717	1 MCCA_MOUSE	Q99mr8 mus musculus
17	891.5	15.2	654	1 BCCA_MYCTU	P46401 mycobacteri
18	875	14.9	448	1 ACCC_BACSU	P49787 bacillus su
19	874.5	14.9	567	1 PYCB_METJA	Q58628 methanococ
20	869.5	14.8	598	1 BCCA_MYLE	P46392 mycobacteri
21	869	14.8	703	1 PCCA_HUMAN	P05165 homo sapien
22	865	14.7	731	1 MCCA_SOYBN	Q42777 glycine max
23	862	14.7	734	1 MCCA_ARATH	Q42523 arabidopsis
24	842.5	14.4	704	1 PCCA_RAT	P14882 rattus norv
25	837.5	14.3	1835	1 DURL_YEAST	P32528 saccharomyc
26	828	14.1	568	1 PYCB_METTH	O27179 methanobact
27	806	13.7	590	1 DCOA_SALTY	Q03030 salmonella
28	801.5	13.7	595	1 DCOA_KLEPN	P13187 klebsiella
29	620.5	10.6	2345	1 COAL_RAT	P11497 rattus norv
30	617.5	10.5	2324	1 COAL_CHICK	P11029 gallus gall
31	617.5	10.5	2346	1 COAL_HUMAN	Q13085 homo sapien
32	616.5	10.5	2346	1 COAL_BOVIN	Q9tt53 bos taurus
33	607.5	10.4	2346	1 COAL_SHEEP	Q28559 ovis aries

## ALIGNMENTS

RESULT 1

ID	PYC_MOUSE	STANDARD;	PRT;	1178 AA.
AC	Q05920:			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Pyruvate carboxylase, mitochondrial precursor (EC 6.4.1.1) (Pyrucv			
DE	carboxylase) (PCB).			
GN	PC OR PCX.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Adipocyte;			
RX	MEDLINE=93189578; PubMed=8446588;			
RA	Zhang J., Xia W.L., Brew K., Ahmad F.;			
RT	"Adipose pyruvate carboxylase: amino acid sequence and domain			
RT	structure deduced from cDNA sequencing."			
RL	Proc. Natl. Acad. Sci. U.S.A. 90:1766-1770(1993).			
CC	-!- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,			
CC	INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY			
CC	ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE			
CC	CARBOXYL GROUP TO PYRUVATE IN THE SECOND. CATALYZES IN A TISSUE			
CC	SPECIFIC MANNER, THE INITIAL REACTIONS OF GLUCOSE (LIVER, KIDNEY)			
CC	AND LIPID (ADIPOSE TISSUE, LIVER, BRAIN) SYNTHESIS FROM PYRUVATE.			
CC	-!- CATALYTIC ACTIVITY: ATP + pyruvate + HCO(3)(-) -> ADP + phosphate +			
CC	oxaloacetate.			
CC	-!- COFACTOR: BIOTIN AND MANGANESE.			
CC	-!- PATHWAY: GLUCONEOGENESIS AND LIPOGENESIS.			
CC	-!- SUBUNIT: HOMOTETRAMER.			
CC	-!- SUBCELLULAR LOCATION: Mitochondrial matrix.			
CC	-!- TISSUE SPECIFICITY: LIVER, KIDNEY, ADIPOSE TISSUE, LIVER			
CC	AND BRAIN.			
CC	-!- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPOAMIDE TRANSFERASES			
CC	AND CARBAMYL PHOSPHATE SYNTHETASES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; L09192; AAA39737.1; -			
DR	PIR; A47255; A47255.			
DR	HSSP; P24182; 1BNC.			
DR	SWISS-2DPAGE; Q05920; MOUSE.			
DR	MGD; MGI:97520; PCX.			
DR	InterPro; IPR001882; PCX.			
DR	InterPro; IPR000089; Biotin_lipoyl.			
DR	InterPro; IPR000901; CPSase.			
DR	InterPro; IPR000891; HMGL-like.			

P32874 saccharomyc  
Q00763 homo sapien  
Q00955 saccharomyc  
P78820 schizosacch  
O28994 archaeglob  
O27077 methanobact  
Q8xa38 escherichia  
Q9wz27 thermotoga  
P00968 escherichia  
P14846 salmonella  
Q82917 salmonella  
P03965 saccharomyc

34 570.5 9.7 2273 1 HFA1\_YEAST  
35 560.5 9.6 2483 1 COA2\_HUMAN  
36 547.5 9.3 2233 1 COAC\_YEAST  
37 519 8.8 2280 1 COAC\_SCHPO  
38 267.5 4.6 1076 1 CARB\_ARCFU  
39 266 4.5 1060 1 CARB\_METTH  
40 255 4.3 1072 1 CARB\_ECO57  
41 254.5 4.3 1099 1 CARB\_THEMEA  
42 252 4.3 1072 1 CARB\_ECOLI  
43 248.5 4.2 1074 1 CARB\_SALTY  
44 245.5 4.2 1074 1 CARB\_SALTY  
45 244 4.2 1118 1 CARB\_YEAST

DR InterPro: IPR003379; PYC\_OADA.  
 DR Pfam: PF00289; Cpsase\_Lchain; 1.  
 DR Pfam: PF00364; biotin\_lipoyl; 1.  
 DR Pfam: PF00682; HMGU-like; 1.  
 DR Pfam: PF02436; PYC\_OADA; 1.  
 DR Pfam: PF02785; Biotin\_carb\_C; 1.  
 DR Pfam: PF02786; Cpsase\_LD2; 1.  
 DR TIGRFAMs: TIGR01235; pyruv\_carbox; 1.  
 DR PROSITE: PS00188; BIOTIN; 1.  
 DR Ligase; Multifunctional enzyme; Biotin; Manganese; Gluconeogenesis;  
 KW ATP-binding; Mitochondrion; Lipid synthesis; Transient peptide.  
 KW TRANSIT 1 20 MITOCHONDRION (POTENTIAL).  
 FT CHAIN 21 1178 PYRUVATE CARBOXYLASE.  
 FT DOMAIN 21 549 BIOTIN CARBOXYLASE (BY SIMILARITY).  
 FT DOMAIN 550 1000 CARBOXYLTRANSFERASE (BY SIMILARITY).  
 FT DOMAIN 1096 1178 BIOTIN CARBOXYL CARRIER PROTEIN  
 (BY SIMILARITY).  
 FT NP\_BIND 198 203 ATP (BY SIMILARITY).  
 FT ACT\_SITE 328 328 BY SIMILARITY.  
 FT BINDING 1144 1144 BIOTIN (BY SIMILARITY).  
 FT SEQUENCE 1178 AA; 129684 MW; 14CEA0F9DA8B8127 CRC64;  
 Query Match 43.5%; Score 2550.5; DB 1; Length 1178;  
 Best Local Similarity 46.1%; Pred. No. 2.5e+129;  
 Matches 543; Conservative 192; Mismatches 407; Indels 35; Gaps 13;  
 7 GGLLLGIIILVSTHTSTLPA-----FKILVANRGEIAVRAFAALETGAA 54  
 9 GGLRLGV-----RRSSAPVSPNRRLEYKPIKVVANRGEIAIRVRACTELGIR 62  
 55 TVAIYPRDRGSRFURSPASEAVRIGTSGPVKAYLDIDEIIGAARKVADAIYPGYGFLS 114  
 63 TVAVISEDTGMHROKADAYLIGRLAPVQAYLHPDIKIVAKENGVDVHPGYGFLS 122  
 115 ENAOLARECAENGITFIGTPEVLDTLGDKSRAVTAAKAGLPVL-AESTPSKNIDIVK 173  
 123 ERAFPAQACQAGVRFTGPSVEVYKMGDKVEARAIAGAIVPVPGTDSPISSLHEAHE 182  
 174 SAEGQTFPIFYKAVAGGGGMRVSPDELRKLKATASREAAAFGDSGVYVERAVINP 233  
 183 FSNFTGFPIIFKAYAGGGGMRVHVSYEELENYTRYASEALAAFGALFVEKFTIEK 242  
 234 QHIEVQILGDRTEGVVHLYERDCSLQRHQKRVVEIAPAOHLDPRLDRICADAVKFCRSI 293  
 243 RHIEVQILGDQYGNILHLYERDCSLQRHQKRVVEIAPATHLPDLRLSDSVKLAKOV 302  
 294 GYQAGATVEFLVDEKGNHVFTEMPRIQVEHTVTEVTEVDLVKAQMLAAGATLKLGL 353  
 303 GYENAGTVEFLVDEKGNHVFTEVNSRLQVEHTVTEITDVLVHAQTHVSEGRSLPDLGL 362  
 354 TQDKIKTHGAALQCRITTEDPNNGFRPDGTITAYRSPGGAGVRLDGAOL-GGEITAHF 412  
 363 RQENIRINGCAICQKRVTTEDPARSFQPDGTGRIEVRSGEGMIRLDNASAFQGAVISPHY 422  
 413 DMLVKMTCRGSDPETAVARAQRLAEFTVSGVATNIGFLRLALLREDEFTSKRIATGFIG 472  
 423 DSLLVKVIHGHKDHPTAATKMSRALAEFRVGRVKTNPFLQNLNNGOFLAGIVDTQFID 482  
 473 DPHLLQAPADDEGRILDYLDVATYKNGHYRPPKDA-APTD-KLPNTKDLPLRGRS 530  
 483 ENPELFQURPAQNRQAKLLHYLGHVWNGTTPPIPVNVSPPDPAPVVPVPGPPAGFR 542  
 531 DRLLQOLGPAFARDLRODALVDTTFRDAHQSLLATRVRSFALPAAVAKLTPPELL 590  
 543 DILLREGPEGFARVNRHQGLLMDTTFROAHOSLLATRVTRDHLKRIADYVAHNKFLF 602  
 591 SVEAGGATYDVAMRFLFEDPMDRLDELREAMPNVNOMLLRGNTVGYTPYPPDSVCRAF 650  
 603 SMENMGATFDVAMRFLFECYPRRLQELRELINIPQMLLRGANAVGTYNPDNVVFKF 662  
 651 VKEAASSGVDFITFDALNDVDSQMRPAIDVLETNTAVAEVAMAYSGDLSDPNEKLYTLTD 710  
 663 CEVAKENGMDVFRVFDLSNLYPNMLLQMEAGSAG-GVVEAAIISYTGVDADPSRTRYSE 721

QY 711 YLKMMAEIVKSGAHLAIKIDMAGLRLPAAVTKLVLTALRREF-DLPVHVHTHTDTAGQOLA 769  
 DB 722 YNGLAEELVRAGTHILCIKIDMAGLRLKPAACMLVSSLRDRFPDLPLHHTHTDTSGAGVA 781  
 QY 770 TYFAAAAGADAVDGSAPLSGTTTSPSLSAIVAAFAHTRRDTGLSLEAVSDLEPYWEAV 829  
 DB 782 AMLACAQAGADVVDVAVDSMSGMTSPQSMGALVACTKGTPLDTEVPLRVFDYSEYWEA 841  
 QY 830 RGLYLPFESGTPGPTGR--VYRHEITPGGQLSNLRAQATGALGLADLADLEEDNYAAVNEML 887  
 DB 842 RGLYAAFDCTATMKSNGSDVYENIEPGGYTNLHFQAHSMGLSGSKFVKKAYVEANQML 901  
 QY 888 GRPTKVTPTSSKVVGDIALHLVGADPADFAADPOKYDIPDSVIAFLRGELGNPPGWPPE 947  
 DB 902 GDILKVTPTSSKIVGDIAQFMVQNGLSRAEAAQAEELSPRSVVEFLOGYIGIPHGGEPE 961  
 QY 948 PLRTRAL-----EGRSECKAPLTVPEPEQAHLDADDSKERRNSLRLPLPKTEEFL 1000  
 DB 962 PFRSKVLKDLPRIEGRPGASLPPLNLKEKLDLIDRHGEVTPEDVLSAAMYPDVFAQPK 1021  
 QY 1001 EHRHRRFGNTSALDDREFFYGLVEGRETILRLPDVTPPLVRLDALSEDDDKGRMVVAVY 1060  
 DB 1022 DTAFTFGPLDSNTRFLFLOQPKIAEEFEVELERGKT-LHIKALAVSDLNLRAGQOVFFEL 1080  
 QY 1061 NGQIRPMRVDRSRVSIVTATAEKADSSNKGHVAAFPAG-VVTVTVTAEGDEGVKAGDAVAII 1119  
 DB 1081 NGOLRSILVKDTQAMKEMHFHFKALKDKVGQIGAPMPGKVIDIKVAAGDKVAKGQPLCVL 1140  
 QY 1120 EAMKMEATITASVDGKIERYVVPVPAATKVEGGDLIVVV 1156  
 DB 1141 SAMKMETVTVSPMEGTIRKVHVTKDMLTLEGDDLILEI 1177  
 RESULT 2  
 PYC\_RAT  
 ID PYC\_RAT STANDARD; PRT: 1178 AA.  
 AC P52873; Q64555;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Pyruvate carboxylase, mitochondrial precursor (EC 6.4.1.1) (Pyruvic  
 DE carboxylase) (PCB).  
 GN PC.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1] SEQUENCE FROM N.A.  
 RP TISSUE=Liver.  
 RX MEDLINE=96096548; PubMed=8522203;  
 RA Lehn D.A., Moran S.M., Macdonald M.J.;  
 RT "The sequence of the rat pyruvate carboxylase-encoding cDNA.";  
 RL Gene 165:331-332(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar;  
 RX MEDLINE=96257760; PubMed=8687410;  
 RA Jitrapakdee S., Booker G.W., Cassidy A.I., Wallace J.C.;  
 RT "Cloning, sequencing and expression of rat liver pyruvate  
 RT carboxylase.";  
 RL Biochem. J. 316:631-637(1996).  
 CC -!- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,  
 CC INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY  
 CC ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE  
 CC CARBOXYL GROUP TO PYRUVATE IN THE SECOND. CATALYZES IN A TISSUE  
 CC SPECIFIC MANNER, THE INITIAL REACTIONS OF GLUCOSE (LIVER, KIDNEY)  
 CC AND LIPID (ADIPOSE TISSUE, LIVER, BRAIN) SYNTHESIS FROM PYRUVATE.  
 CC -!- CATALYTIC ACTIVITY: ATP + pyruvate + HCO(3)(-) = ADP + phosphate +  
 CC oxaloacetate.  
 CC -!- COFACTOR: BIOTIN AND MANGANESE (BY SIMILARITY).  
 CC -!- PATHWAY: GLUCONEOGENESIS AND LIPOGENESIS.

Qy	354	TQDKIETHGAALOCRTITTEDPNNGFRFPORGTITAYRSPGGAGVRLDGAQAL-GGEITAHF	412
Db	363	ROENIRINGCAICROCVTTIEDPARSFOPDFTGRIEVRSEGGMGIRLDNLSAFQGAVISPY	422
Qy	413	DSMLVMTKCRGSGFETAVARAORALAEFTVSGVATNIGFLRALLREEDFTSKRTATFGI	472
Db	423	DSLLVKVIAGKDHPTAATYKMSALAEFRVGVKYNIPQLWNQOFLAGIVDTQPID	482
Qy	473	DHPHLQAPPADDEQGRILDYLDVTVNKPHGVRPKDVA-APIDKL-PNIKDLPLPRGSR	530
Db	483	ENPELFQLRPAQNRAKQLLHYLGHVVMVNGPTTPIPVKVSPPSDPIVPWPVTGPPAGFR	542

QY	351	DRKJQJPARAFKRLREQJALAVDITFI	KDAHQSLUATIKVRSFALKPAAEAVAKLTPELL	390
Db	543	DILLREGCEGFARAVNRHOGLLIMDTT	FDFAHQSLLATRVTRHDLKKTAPYVAHNFNMLF	602
QY	591	SVEAWGATFDVAMRLEFEDPDWRDLDEL	EAMPNVNIQMLLRGNTVGYTYPDVSVCRAF	650
Db	603	SIENWGATFDVAMRLEYECPRHRRLOEL	RELIPNIPFQMLLRGANAVGTYNPDVYKPF	662
QY	651	VKEAASGVGDIIFRIDALNDVSOMRPAID	AVLETNTAAEAVAMAYSGDSDPNKELYTLD	711
Db	663	CEVAKENGMDVFRIFDLSNLPNMLLMGE	AAGSAG-GVVEAAISYTGVDADSPRTKYSLE	721
QY	711	LYLKMAEEIVKSGAHTLAKDMAGLLRPA	AVTKLYTALRRF-DLPVHVHHTHDTAGGOLA	769
Db	722	YYNGLAPELVRAGTHILCKDMAGLLKPA	ACTMLVSSURDRFPDLPJLHIHTHDTFSGSVA	781
QY	770	TYFAAAQAGADAVDGAAPLSGTTTSOP	LSUSATVAFAHTRRDTGLSLEAVSDLPSEAV	829
Db	782	AMLACAQAGADVDAVDSKSGMTSOPSM	GALVACTKGTPLDTEVLERVDFSEYHEGA	841
QY	830	RGLYLPFESGTPGPTGR--VYRHEITP	GGOLSNLRQAATGALADRFELIEDNYAAVNML	887
Db	842	RGLYAAFDCTATMKSGNSDVIYENED	PGGYTNLHFQAHSMGLGSFKCVKAYVEANQML	901
QY	888	GRPTKYTPSSKVVGDIALHLVAGVDPA	DFADPKYDIPQSVIAFLRGELNPGGWPE	947
Db	902	GDLIKYTPPSKIVGDIAQFWMVONGUS	RABEAQAEELSFPRSVEFLOGYIGIPHGGFPE	961
QY	948	PLRTRALEG--RSEGGAPLTEVPEEQAH	LADAD-----DSKERNNSLRLLPKPTPEEF	999
Db	962	PFRSKVLKDLPRLEG-GPGASLPLNLK	LEKXDLDRHGEVTPEDVLSAAMYPDVFAOF	1020

Db	1021	KDFTATFGPLDSLNRFLQGP	KIAEEFEVELERCKT-LHIKALAVS	DLNRAGQ	QVFFE	1079
Qy	1060	VNGQIRPMVRDRSVESVTATA	EKAEDSSNKGHVA	APFAG-VVTVTV	VAEGDEVKAGDAV	1118
Db	1080	LNGLLSILVKDTQAMKEMHF	HPKALKDVKGQIGAP	MPGKVIDVKVAAGAKV	VWGPCLV	1139
Qy	1119	IEAMKMEATITASVDGKI	ERVVVPAATKVEGG	DLIVV	1156	
Db	1140	LSAMKETVVVTSPEG	TIRKVVHTKMTLEG	DDLILEI	1177	
RESULT 3						
PYC_HUMAN						
ID	PYC_HUMAN	STANDARD:	PRT:		1178 AA.	
AC	P11498:	Q16705:				
DT	01-OCT-1989	(Rel. 12, Created)				
DT	01-OCT-1996	(Rel. 34, Last sequence update)				
DT	15-JUN-2002	(Rel. 41, Last annotation update)				
DE	Pyruvate carboxylase, mitochondrial precursor (EC 6.4.1.1) (Pyruvic					
DE	carboxylase) (PCB).					
GN	PC.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
OX	NCBI_Taxid-9606;					
RP	[1]					
RP	SEQUENCE FROM N.A.					



```
Db 363 RQENIRINGCAIOCRVTTEDPARSFQDPTGRIEVRSGEGGIRLDNASAFQGAVIDSPHY 422
QY 413 DMLVKMTCRSGDEETAVARAQALAEFTVSGVATNIGFLRALLREEDFTSKRIATGFIG 472
Db 423 DLSLVKVIANGKHDTPTAKMSALAEFRVGKNTIAFLONVNNQOFLAGTVDTQFID 482
QY 473 DHPHLLQAPPADDOGRILDYLDADVTVNKPVGVRP-KDVAAPIDK-LPNIKDLPPLRGRSR 530
Db 483 ENPELFOLRPAQNAQKLLHLYGHVMVNGTTPIPYKASPSPTDPVVPVPIGPPAGFR 542
QY 531 DLRLQGLGPAAPARLDLQDQDALAVDTTFRDAHOSLLATRVSRFALKPAABEAVAKLTPELL 590
Db 543 DILLREGPEGFARAVRNHPHGLLLMDTTFRDAHOSLLATRVTHDLKKIAPVAHNEFSKLF 602
QY 591 SVAWGATGVDMARFLFEDPWDLDELREAMPNVNQLMRLGRNTVGYTPYDSDVCRAF 650
Db 603 SMENMGATFDVAMRFLEYECFWRRLQELRIENIPFQMLLRGANAVGYTNYDNNVYKFE 662
QY 651 VKAAASGVDFIRFIDALNDVSQMRPAIDAVLETNTAVAEMAYSGDLSDPNKLYTLD 710
Db 663 CEVAKENGMDVERFDSLNLPLNMLGMEAGSAG-GVVEAAISYTGVDVADPRTKYSIQ 721
QY 711 YYLKMAEEIVKSGAHILAIDKMAGLRPAATVTLVTLALREF-DLPVHVHTHTAGQOLA 769
Db 722 YYMGLAEELVRAGTHILCIKMAGLLKPTACTMLVSLSLDRFPDPLPHIHTDTSAGVA 781
QY 770 TYFAAAQAGADAVDGAAPLSGTTSPSLSAIIVAFAHTTRDTGLSLEAVSDLEPTWEAV 829
Db 782 AMLACAQAGADVVDVAADSMGSMGTSQPSMGALVACTRGTPDLDTEVPMERVDYSEWEGA 841
QY 830 RGLYLPFESGTPGPTGR--VYRHEIPGQSLNLRQAATLGLADRLADRFELIEDNYAAVNEML 887
Db 842 RGLYAFDCTATKSGNSDYENEIPGQVYTNLFQAHSMGLSKGKEVKKAIYEAQNML 901
QY 888 GRTKVTYPSKVYGDIALHLVAGVDPADPAADPQKYDIPDSVIAFLRGLGNPPGQWPE 947
Db 902 GDLIKVTPSSKIYGDIAQFVQVQNGLSRAEAAQAEELSPRSVVEFLQYIGVPHGQFPE 961
QY 948 PLRTRAL-----EGRSKAGPLATEVPEERQAHLDL-DOSKERRNSLNRLFFPKPTEEFL 1000
Db 962 PFRSKVLKDLPRVEGRGASLPDLIDQALEKELVDHRGEEVTPEDVLISAAMYPDVFAHEK 1021
QY 1001 EHRRFRGNTSALDDREFFYGLVREGRETILRLPDVTRPLLVRLDAISEPDDKGMNRNVAVV 1060
Db 1022 DFTATGGLDLSNTRFLQPKTAEFEVEFLERKGT-LHIKALVSDLNRAQORQVFEFL 1080
QY 1061 NGQIRPMRVDRVSVESTAEKADSSNKGHVAAFPAG-VVTVVTAEGDEVKAGDAVAII 1119
Db 1081 NGQLRSILVKDQAMKEMHFHPKALDKVKGQIGAPMPGKVIDIKVVAGAKVAKGQPLCVL 1140
QY 1120 EAMKMEATITASVDGKTIERVVVVPAATKVGGDLIVVV 1156
Db 1141 SAMKMETVTSPEMGTIVRKVVHVTKMDTLEGDDLILEI 1177

RESULT 4
PYC1_YEAST
ID PYC1_YEAST STANDARD; PRT; 1178 AA.
AC P11154;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pyruvate carboxylase 1 (EC 6.4.1.1) (Pyruvic carboxylase 1) (PCB 1).
GN PYC1 OR PYV OR YGL062W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=88298805; PubMed=3042770;
RA Lim F., Morris C.P., Occhiodoro F., Wallace J.C.;
```

```
RT "sequence and domain structure of yeast pyruvate carboxylase.";
RL J. Biol. Chem. 263:11493-11497(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX Feuermann M., de Montigny J., Potier S., Souciet J.-L.;
RA "The characterization of two new clusters of duplicated genes
RT suggests a 'leugo' organization of the yeast Saccharomyces cerevisiae
RT chromosomes.";
RL Yeast 13:861-869(1997).
RN [3]
RP SEQUENCE OF 1003-1178 FROM N.A.
RX MEDLINE=87241529; PubMed=3036126;
RA Morris C.P., Lim F., Wallace J.C.;
RT "Yeast pyruvate carboxylase: gene isolation.";
RL Biochem. Biophys. Res. Commun. 145:390-396(1987).
CC -!- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,
CC INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY
CC ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE
CC CARBOXYL GROUP TO PYRUVATE IN THE SECOND.
CC -!- CATALYTIC ACTIVITY: ATP + pyruvate + HCO(3)(-) = ADP + phosphate +
CC oxaloacetate.
CC -!- COFACTOR: BIOTIN AND ZINC.
CC -!- PATHWAY: GLUCONEOGENESIS.
CC -!- SUBUNIT: HOMOTETRAMER.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPOAMIDE TRANSFERASES
CC AND CARBAMYL PHOSPHATE SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J03889; AAA34843.1;
CC EMBL: Z72584; CAA96765.1;
CC PIR: A29233; QYBYP.
CC HSSP: P24182; 1BNC.
CC SGD: S0003030; PYC1.
CC InterPro: IPR001882; Biotin_attach.
CC InterPro: IPR000089; Biotin_lipoyl.
CC InterPro: IPR000901; CPSase.
CC InterPro: IPR000891; HMGL-like.
CC InterPro: IPR003379; PYC_OADA.
CC Pfam: PF00289; CPSase_L_chain; 1.
CC Pfam: PF00364; biotin_lipoyl; 1.
CC Pfam: PF00682; HMGL-like; 1.
CC Pfam: PF02436; PYC_OADA; 1.
CC Pfam: PF02785; Biotin_carb_C; 1.
CC Pfam: PF02786; CPSase_L_D2; 1.
CC TIGRFAMS: TIGR01235; pyruv_carbox; 1.
CC PROSITE: PS00188; BIOTIN; 1.
CC PROSITE: PS00866; CPSASE_1; 1.
CC PROSITE: PS00867; CPSASE_2; 1.
CC Ligase; Multifunctional enzyme; Biotin; Gluconeogenesis; ATP-binding;
KW Zinc; Multigene family.
FT NP_BIND 182 187 ATP (POTENTIAL).
FT ACT_SITE 312 312 BY SIMILARITY.
FT BINDING 1135 1135 BIOTIN (BY SIMILARITY).
FT SIMILAR 160 330 CARBAMOYL PHOSPHATE SYNTHETASES.
FT SIMILAR 350 470 WITH OTHER BIOTIN CARBOXYLASES.
FT SIMILAR 1086 1178 WITH OTHER BIOTIN CARRIER PROTEINS AND
FT WITH LIPOAMIDE ACETYLTTRANSFERASE.
FT CONFLICT 462 462 T -> G (IN REF. 1).
FT CONFLICT 493 493 V -> D (IN REF. 1).
FT CONFLICT 595 595 R -> A (IN REF. 1).
FT CONFLICT 619 619 E -> Q (IN REF. 1).
FT CONFLICT 664 664 G -> S (IN REF. 1).
FT CONFLICT 772 772 A -> R (IN REF. 1).
```

QY 972 QAHL DADDSKERRNSLNL-----LFPKPTEEEFLEHRRRFGNTSALDDREFFY 1019

DR EMBL; X59890; CAA42544.1; -.  
 DR EMBL; Z36087; CAA85182.1; -.  
 DR EMBL; U35647; AAC49147.1; -.  
 DR PIR; S46094; S46094.  
 DR HSP; P24182; IBNC.  
 DR SGD; S0000422; PYC2.  
 DR InterPro; IPR001882; Biotin\_attach.  
 DR InterPro; IPR000089; Biotin\_lipoyl.  
 DR InterPro; IPR000901; CPSase.  
 DR InterPro; IPR000891; HMG-L-like.  
 DR InterPro; IPR003379; PYC\_OADA.  
 DR Pfam; PF00289; CPSase\_L\_chain; 1.  
 DR Pfam; PF00364; biotin\_lipoyl; 1.  
 DR Pfam; PF00682; HMG-L-like; 1.  
 DR Pfam; PF02436; PYC\_OADA; 1.  
 DR Pfam; PF02785; Biotin\_carb\_C; 1.  
 DR Pfam; PF02786; CPSase\_L\_D2; 1.  
 DR TIGRFAMS; TIGR01235; pyruv\_carbox; 1.  
 DR PROSITE; PS00188; BIOTIN; 1.  
 DR PROSITE; PS00866; CPSASE.1; 1.  
 DR PROSITE; PS00867; CPSASE.2; 1.  
 KW Ligase; Multifunctional enzyme; Biotin; Gluconeogenesis; ATP-binding;  
 KW Zinc; Multigene family.  
 FT NP\_BIND 183 188  
 FT ACT\_SITE 313 313  
 FT BINDING 1136 1136  
 FT SIMILAR 161 331  
 FT SIMILAR 351 471  
 FT SIMILAR 1087 1180  
 FT CONFLICT 15 15  
 FT CONFLICT 132 132  
 FT CONFLICT 238 238  
 FT CONFLICT 268 268  
 FT CONFLICT 546 546  
 FT CONFLICT 642 642  
 FT CONFLICT 771 773  
 FT CONFLICT 831 831  
 FT CONFLICT 839 839  
 FT CONFLICT 1001 1001  
 FT CONFLICT 1155 1155  
 FT CONFLICT 1178 1178  
 FT CONFLICT 1180 1180  
 FT SEQUENCE 1180 AA; 130166 MW; AD60DA3A60F5E001 CRC64;

Query Match 41.5%; Score 2436; DB 1; Length 1180;  
 Best Local Similarity 45.5%; Pred. No. 3.5e-123;  
 Matches 530; Conservative 185; Mismatches 407; Indels 44; Gaps 16;

QY 24 STLPAFKKILVANRGEIAVRAFAALLETGAATVAIYPREDRGSFHRSPASEAVRIGTEG- 82  
 DB 15 SLLGEKNKILVANRGEIPIRISAEHLSMRTIAIYSHEDRLSMHRLKADAEYVIGEEQ 74

QY 83 -SPVRAKLDIDETIGAAKKVADAIYPGVGFLSENQAOLRECAENGITIGTPEVLDLT 141  
 DB 75 YTPVGAYLAMDEIIETIAIKHKVDIFHPGVGFLSENSEFADKVKAGITWIGPPAEIVDSV 134

QY 142 GKSRAVTAAKAGLPVLAEST-PSKNIDDIKSAEGQYPIFVKAVAGGGGRGMRVSS 200  
 DB 135 GDKVSARHIAAANRVPVPTGPIETVOEALDFVNEYGYPIIIRKAAFGGGGRGMRVRE 194

QY 201 PDELRRKATEASREAAAFGDGSGVYVERAVINPQHIEVQILGDRTEGVVHLVERDCSLQR 260  
 DB 195 GDOVADAFORATSEARTAFNGTCFVERFLDKPKHIEVOLLDADNHNHGVHLPERDCSVQR 254

QY 261 RHQKVVEIAPAOHLDPDLDRIDCADAVKFCRSIGYOGAGTVEFLVDEKGNHVFIEINPRI 320  
 DB 255 RHQKVVEVAPAKTLPREVRDAILTDVAKLAKVCGYRAGTAETFLVDNQNRHVFIEINPRI 314

QY 321 QVEHVTVEEVDVLKQAMRLAAGATLKLGLTQDKIKTHGAALOCRITTEDPNNGFPP 380  
 DB 315 QVEHTITEITGIDIVSAQIQIAAGATLQGLLQDKITTRGFSIQCRITTEDPSKNFQP 374

QY 381 DTGTTAYRSPGGAGVRLDGA-AQLGGEITAHFDSMLVKMTCRGSDFETAVARAQALAE 439  
 DB 375 DTGRLEVYSAGNGVRLDGGNAYAGATISPHYDSMLVKCSGSIYEVIRRMIRALIE 434  
 QY 440 FTVSGVATNIGFLRALLREEDFTSKRIATGFIGDHPHLLQAPADDEQGRDILYLDADV 499  
 DB 435 FRIRGVKTNIPFELLTLLTNPVFTGTYWTFIDDTQOLFQWSSNQAKLHLHYLDLAV 494  
 QY 500 N---KPHGVKRPDVAAPIDKLPNTKDL-----PLPRGSRDLKOLGPAFAARDL 545  
 DB 495 NGSSIKGOIGLPLKLSNP--SVPHLHDAOGNINVTKSPAPSGWROVLLEKGPSEFAK 552  
 QY 546 REQDALAVTDTTFRDAHQSLATRVRSFALKPAEAAVAKLTPELLSVEAGGATYDAMR 605  
 DB 553 RQFNGLTLLMDTTWROAHQSLATRVTHDLATITPTTAHALAGAFACGCGATFQVAMR 612  
 QY 606 FLFEDPWRDLDELREAMPNVNTOMLGRNTVGYTPYDSVCFRAFYKEAASSGVDFIRF 655  
 DB 613 FLHEDPWERLRLKRLSLVPNIPFQMLLRGANGVAYSSLPDNDHFVKQAKDNGVDIFRVF 672  
 QY 666 DALNDVSOMRPAIDAVLETNTAVAEVAMAYSGDLSDPNEKLYTLDYYLKMAREIVKSGAH 725  
 DB 673 DALNDELUKVGVNAVKKAG-GVVEATVCGSGMDLQPGKK-YNLDIYLEVKEIVONGTH 730  
 QY 726 ILAIKDMAGLLRPAAVTKLVTLALREF-DLPVHVHTHTDAGGQATYFAAAQADAVDG 784  
 DB 731 ILGKIDMAGTMKPAAKLLIGSLRTRYPDLPIHVHSDSAGTAVASMTACALAGADVVDV 790  
 QY 785 ASAPISGTTQSQSLSAIIVAAFAHTRDRTGLSLEAVSDLEPYWEAVRGLYLPFESGTPGPT 844  
 DB 791 AINSMSGLTSQPSINALLASL-EGNIDTGTINVEHVRELDAYNAEMRLLYSCFEADLKGP 849  
 QY 845 GRVYRHEIPGGOLSNLRAOATLGLADRFELLEDNYAAVNEMLGRPTKPTSSKVVGDILA 904  
 DB 850 PEYVTHEIPGGQTLNLLFQAQQLGEGEQAETRAVREANYLUGDVIKVTPTSKVVGDLA 909  
 QY 905 LHLVAGVDPADFAADPQKYDIPDSVIAFLRGLGNPGGWPPLRTRALEGRSEKAPL 964  
 DB 910 QFMVSNKLTSDDIRRLANSLLDFPDSVMDFFEGILGQYGFPEPLSLDLRNK---RRKL 966  
 QY 965 TEVPEEEQHLDAADSKERRNSLNL-----LFPKTEFELEHRRRFGNTSAL 1012  
 DB 967 TCRPGLEPFPDLE--KIREDLQNREFGIDECDVASNNYPRVYEDFQKIRETYGDLVSL 1024  
 QY 1013 DDERFPYGLVEGRETLIRLPDVRTPLLRDLDAISEPDDK-GMRNVVANVNGQIRPMVRD 1071  
 DB 1025 PTKNFILAPPEDEIEVTEEQGT-LIILQAVGDLNKKTGQREVYFELNGELRKRIVAD 1083  
 QY 1072 RSVESVTATAEKADSSNKGHVAAPFAGV-VTVTVAEGDEVKAGDAVAIIEAMKMEATITA 1130  
 DB 1084 KSONIGSVAKPRADVHTHQIGAPMAGVITIEVKHKGSLVKKGESIAVLSAMKMEMVSS 1143  
 QY 1131 SVDGKIERVVVPAATVKEGGDLIVV 1156  
 DB 1144 PADGQVKDFIKDGESVDASDLVLV 1169

## RESULT 6

PVC\_PICPA

ID PVC\_PICPA

AC P78992;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Pyruvate carboxylase (EC 6.4.1.1) (Pyruvic carboxylase) (PCB).

GN PYC1.

OS Pichia pastoris (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Pichia.

OX NCBI\_TaxID=4922;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98301182; PubMed=9639311;



RA Menendez J., Delgado J., Gancedo C.:  
 RT "Isolation of the *Pichia pastoris* pyC1 gene encoding pyruvate  
 RT carboxylase and identification of a suppressor of the pyC  
 RT phenotype".  
 RL Yeast 14:647-654(1998).  
 CC -1- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,  
 CC INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY  
 CC ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE  
 CC CARBOXYL GROUP TO PYRUVATE IN THE SECOND (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + pyruvate + HCO(3)(-) -> ADP + phosphate +  
 CC oxaloacetate.  
 CC -1- COFACTOR: BIOTIN AND ZINC.  
 CC -1- PATHWAY: GLUCONEOGENESIS.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPOAMIDE TRANSFERASES  
 CC AND CARBAMYL PHOSPHATE SYNTHETASES.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Y11106; CAA71993.1; -  
 DR HSSP: P24182; LDV1.  
 DR InterPro: IPR001882; Biotin\_attach.  
 DR InterPro: IPR000089; Biotin\_lipoyl.  
 DR InterPro: IPR000901; CPSase.  
 DR InterPro: IPR000801; HMGU-like.  
 DR InterPro: IPR003379; PYC\_OADA.  
 DR Pfam: PF00289; CPSase\_L\_chain; 1.  
 DR Pfam: PF00364; biotin\_lipoyl; 1.  
 DR Pfam: PF00682; HMGU-like; 1.  
 DR Pfam: PF02436; PYC\_OADA; 1.  
 DR Pfam: PF02785; Biotin\_carb\_C; 1.  
 DR Pfam: PF02786; CPSase\_L\_D2; 1.  
 DR TIGRfams: TIGR01235; pyruv\_carbox; 1.  
 DR PROSITE: PS00188; Biotin; 1.  
 DR PROSITE: PS00866; CPSASE\_1; 1.  
 DR PROSITE: PS00867; CPSASE\_2; 1.  
 DR KW Ligase; Multifunctional enzyme; Biotin; Gluconeogenesis; ATP-binding;  
 DR Zinc.  
 FT NP\_BIND 185 190 ATP (POTENTIAL).  
 FT ACT\_SITE 315 315 BY SIMILARITY.  
 FT BINDING 1140 1140 BIOTIN (BY SIMILARITY).  
 SQ SEQUENCE 1189 AA; 131400 MW; 8B6E858079657914 CRC64;

Query Match 40.8%; Score 2393; DB 1; Length 1189;  
 Best Local Similarity 45.5%; Pred. No. 7 3e-121;  
 Matches 529; Conservative 176; Mismatches 419; Indels 38; Gaps 18;

QY 23 SSTLPFAKKILVANKGEIIVAFRAALETGAATVAIYPREDGSGFHRSPASEAVRIGTEG 82  
 DB 16 SSLGCTMKNKILVANKGEIPIRTIETHLSMNTVAIYSHEDRLSMRLKADEAYVIGERG 75  
 QY 83 --SPVKAYLDIDEIGAAKKVADAIYGVGFTLSENAQLARECAENGITFTGTPPEVLDL 140  
 DB 76 QYSPVQAYLAIDEIKIYAVKHNNMTHPGYGFCSENSEFAKVEENGILLWVGSDTVIDA 135  
 QY 141 TGKSRVATTAAGKAGLPVLAESTPSKNIDIVKS---AEGQTYPIPVKAVAGGGRGMRP 197  
 DB 136 VGDKVSARNLAYAANVPTV-PGTPGP-IEDYAQATAFVEEYGYPTVIKAAFGGGGRGMRV 193  
 QY 198 VSSPDLKRLKATEASRAEAFAAGSGSVYVERAVINPQHIEVQILGDRGTGEVHLYERDCS 257  
 DB 194 VREGDDIEDAFLRASSEAKAFNGTGVFERFLDKPKHIEVQLLDADNYGNVILHLERDCS 253  
 QY 258 LORRHQKVEIAPQHLDPRLRCADAVKFCRSYGCAGTVEFLVDEKGNHVEIEN 317  
 DB 254 VQRHQKVARNCSAKTLPVVERNAILLDAVKLAKTANYRNAGTAEFLVDSQNRHYEIN 313

QY 318 PRIQVEHTVTBEVTEVDLVKQAMRLAAGATLKELGTLQDKIKTHGAALQCRITTEDPNNG 377  
 DB 314 PRIQVEHTITEITGVDIVAAQIQIQAAGASLEQLGLLOEKITTRGFAIOCRITTEDPTKN 373  
 QY 378 FRPDGTGTTAYRSGGAGVRLDGAALQGE-ITAFHDSMLVKMTCRGSDFFATAVARA 436  
 DB 374 FOPDTGKTEVYRSSGNGVRLDGGNGFAGAVLSHYDSMLVKCSTSGSNYEIRRRKMIRA 433  
 QY 437 LAEFTVSGVARNIGFLRALLREEDFTSKRIATGFTGDHPHLLOAPPADDEOGRILLYAD 496  
 DB 434 LVEFRIRGVKNINIFELLALLTHPVMTSECTWTFIDDTPELFKILITSONRAQKLAYLGD 493  
 QY 497 VTVN-----KPHGVPRKQVAAIDKLPNIKD--LPLPRGSRDRRLQQLGPAAPAR 543  
 DB 494 LAVNGSSIKGOIGLPLKLH--READIPSITDINGVDIVSIPPDGWRQFLLEKPEQFAQ 551  
 QY 544 DLREODALAVTDTTFRDAHQSLLATRVRSFALKPAAEAAVAKLTPELLSVEANGGATYDVA 603  
 DB 552 QVRAFPGLMINDTWRDRAHQSLLATRVTHDLLNIAPATSYALHHAFALECMGGATEDVS 611  
 QY 604 MRFLEDPWDLDELREAMPNNVNTQMLLRGNRTVGYTPYDPSVCRAFVKEAASSGVDIR 663  
 DB 612 MRFLEDPWDLRLKLRKAVPNTPFMSLLRGNGVAYYSLPDNAIDHFLKQAKDTGVDYFR 671  
 QY 664 IFDALNDVSONRPAIDAVLETNTAAVEAMAYSGDLSDPNEKLYTLDYYLKMAREIVKSG 723  
 DB 672 VFDALNDIEQLKVGVDVAVKKAG-GVVEATMCSGDMKPKKK-YNLEYVINLATEIVEMG 729  
 QY 724 AHILAIKDMAGLLRPAAYTKLVALTALRREF-DLPVHVHTHTAGGOLATYFAAAAGADAV 782  
 DB 730 THILAVKDMAGTLKPTAAKQLISALRRKFPSLPINHVTDSAGTCVASMVACARADV 789  
 QY 783 DGASAPLSGTTSPQSLSAIVAFAHTRDRDTGLSLEAVSDLEPYEAVRGLYLPESGTPG 842  
 DB 790 TVRVNSMGMTSQPSMSAFIASL-DGEIETGIPENAREIDAYWAEMRLLSCEFDLKG 848  
 QY 843 PGRVYRHEIPGGQSLNRAQATALGLADRFELIBDNVAANVMGLRPTKVTTPSKVYGD 902  
 DB 849 PDPEVYQHEIPGGQTLNLLFQAQVGLGEKWKVETKAYEAANRLGDLVKVTPTSKVYGD 908  
 QY 903 LALHLVGAGVDPADPAADPQKYDIPDSVIAFLRGLNPPGPGWPEPLTRALEGRSE--- 959  
 DB 909 LAQFVMSNKLSEDEYERLASLDFPDSVLDFFEGLMGTPYGGFPEPLRTNVIISGKRRLT 968  
 QY 960 GKAPLTPVEEBEQA---HLDADDKSKERNSL-NRLLFPKPTFEFLEHRRRFGNTSALDDR 1015  
 DB 969 SRPGLTLEPYNIPALREDLEAFRSKVTENDVASYNNYPKVYEAKKQQLXGDLVLPTR 1028  
 QY 1016 EPFY--GLVEGRETLIRLPDVRTPLLRDLDAISE--PDDKGMNVVNVANGQIRPMRVDR 1072  
 DB 1029 NPLSPPKIDEERHVTIVTETRTKTLIIKMAEGELSSQSGTREVYFELNGEMKRVTVEDK 1088  
 QY 1073 SVESVTAEAKADSNKGHVAPFAGVVT-VTVASGDEVKAGDAVAITEAMKMEATITAS 1131  
 DB 1089 NGAVETIIRPKADAINPNEIGAPMAGVVVEVRYHENGVEKKGDPVIAVLSAMKMEVISP 1148  
 QY 1132 VDGKTIERVVPAATKVEGGDLI 1153  
 DB 1149 VAGRIQGTAVKENDSVDSADLI 1170  
 RESULT 7  
 PYCA\_METJA STANDARD; PRT; 501 AA.  
 ID PYCA\_METJA AC Q58626;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Pyruvate carboxylase subunit A (EC 6.4.1.1) (Pyruvic carboxylase A).  
 GN PYCA OR M1229.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcales; Methanocaldococcus.





CC -1- SIMILARITY: TO OTHER BIOTIN-DEPENDENT ENZYMES AND CARBAMOYL-  
 CC PHOSPHATE SYNTHETASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: L14862; AAB51770.1; -;  
 DR HSSP: AP003584; BAB72896.1; -;  
 DR InterPro: IPR004549; ACCC.  
 DR InterPro: IPR000901; CPSase.  
 DR Pfam: PF00289; CPSase\_L\_chain; 1.  
 DR Pfam: PF02785; Biotin\_carb\_C; 1.  
 DR Pfam: PF02786; CPSase\_L\_D2; 1.  
 DR TIGRFAMS: TIGR00514; accc; 1.  
 DR PROSITE: PS00866; CPSASE.1; 1.  
 DR PROSITE: PS00867; CPSASE.2; 1.  
 DR Fatty acid biosynthesis; Ligase; Biotin; ATP-binding;  
 KW Complete proteome. 168 ATP (BY SIMILARITY).  
 FT NP\_BIND 163 293 BY SIMILARITY.  
 FT ACT\_SITE 293 293  
 SQ SEQUENCE 447 AA; 49104 MW; 8A541B38B39E00F9 CRC64;

Query Match 17.6%; Score 1032.5; DB 1; Length 447;  
 Best Local Similarity 48.4%; Pred. No. 1.8e-48;  
 Matches 216; Conservative 71; Mismatches 148; Indels 13; Gaps 4;

QY 29 FKKILVANGETAVRAAFRAALTEGAATVAIYPREDRGSFHRSFASAVRIGTGGSPVKAY 88  
 DB 3 FDKILIANRGEIATLILRACEEMGIATVSHVTDNRALHVLADEAVCIIG-EPASAKSY 61  
 QY 89 LDIDEITGAARKVKADAIYPGYGFLSENALARECAENGITFIGTPTEVDLDTGKSRV 148  
 DB 62 LNIPIITAAALTRNSAIHPGYGFLSENAPFAECADPHHIAFGTPPEALRLMGDKSTAK 121  
 QY 149 TAARKAGLPVLAES-----TPSKNIDIVKSAEQTYPIFKVAVAGGGGRMFVSPDE 203  
 DB 122 ETMOKAGVPTVPVSGEVLVTEQGLE----LAKDIGYVPMIKATAGGGGRMLVRSPE 177  
 QY 204 LRKLATASREAAAFDGSVYVERAVINPOHIEVQILGDRTEGVHLYERDCSLQRRHQ 263  
 DB 178 FVKFLAAQAGAGAGAGNAGVYIEKFTERPRIEFQILADNYGNV IHLGERDCSIQRRNQ 237  
 QY 264 KVEIAPQAHLDPRLDRICADAVKFKRSIGYQAGTVEFLVDEKGNHVIEMNPRIOVE 323  
 DB 238 KLEEAPSPALDRLREKMGQAAVKAQAQFINYTCAGTIEFLDRSGQFYEMENNTRIOVE 297  
 QY 324 HTVTEEVTEVDLVAQMLAAGATLKLGLTQDKIKTHGAALOCRIITDPNNGFRPDTG 383  
 DB 298 HPVTVMYTGVDLLVEQIRIAGGERLR--LTQDQVVLRLGHAIECRINAEDPDHDFRAPG 354  
 QY 384 TITAYRSPGGAGVLDGAALGGEITAHFDSMLVKMTCRGSDEPETAVARAORALAEFTVS 443  
 DB 355 RISGLPVGPGVGRIDSHVYDQIIPYIDSLKGLKLVWGPDRATALNNMKRALRECAIT 414  
 QY 444 GVATNIGFLRALLREEDFTSKRIATGFI 471  
 DB 415 GLPTTIGFHQIMENPQFLOGNVSTSPF 442

RESULT 9  
 PYCA\_ARCFU STANDARD; PRT; 506 AA.  
 ID PYCA\_ARCFU  
 AC O30019;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Pyruvate carboxylase subunit A (EC 6.4.1.1) (Pyruvic carboxylase A).

GN PYCA OR AF0220.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
 OC Archaeoglobaceae; Archaeoglobus.  
 NCBI\_taxid=2234;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475;  
 RA Klenk H.-P., Clayton R.J., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.A., Graham D.E., Kyrpides N.C.,  
 RA Richardson D.L., Kerlavage A.R., Lee N.H., Sutton G.G., Gill S.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Adams M.D., Loftus B.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Badger J.H., Glodek A., Zhou L.,  
 RA Peterson S., Reich C.I., McNeil L.K., Weidman J.F., McDonald L., Utterback T.,  
 RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,  
 RA Sadov P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,  
 RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,  
 RA Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus.";  
 RL Nature 390:364-370(1997).  
 CC -1- FUNCTION: PYRUVATE CARBOXYLASE CATALYZES A 2-STEP REACTION,  
 CC INVOLVING THE ATP-DEPENDENT CARBOXYLATION OF THE COVALENTLY  
 CC ATTACHED BIOTIN IN THE FIRST STEP AND THE TRANSFER OF THE CARBOXYL  
 CC GROUP TO PYRUVATE IN THE SECOND.  
 CC -1- CATALYTIC ACTIVITY: ATP + pyruvate + HCO(3)(-) = ADP + phosphate +  
 CC oxaloacetate  
 CC -1- COFACTOR: ATP, MAGNESIUM (OR MANGANESE OR COBALT), PYRUVATE AND  
 CC BICARBONATE (BY SIMILARITY).  
 CC -1- PATHWAY: GLUCONEOGENESIS.  
 CC -1- SUBUNIT: HETEROCTAMER OF FOUR A AND FOUR B SUBUNITS (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: WITH OTHER BIOTIN CARBOXYLASES, LIPOAMIDE TRANSFERASES  
 CC AND CARBAMYL PHOSPHATE SYNTHETASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AE001090; AAB91012.1; -;  
 DR HSSP: P24182; IBNC.  
 DR TIGR: AF0220; -;  
 DR InterPro: IPR004549; ACCC.  
 DR InterPro: IPR000901; CPSase.  
 DR Pfam: PF00289; CPSase\_L\_chain; 1.  
 DR Pfam: PF02785; Biotin\_carb\_C; 1.  
 DR Pfam: PF02786; CPSase\_L\_D2; 1.  
 DR TIGRFAMS: TIGR00514; accc; 1.  
 DR PROSITE: PS00866; CPSASE.1; 1.  
 DR PROSITE: PS00867; CPSASE.2; FALSE\_NEG  
 DR Ligase; Multifunctional enzyme; Gluconeogenesis; Magnesium; Pyruvate;  
 KW ATP-binding; Complete proteome.  
 FT NP\_BIND 161 250  
 FT ACT\_SITE 230 250 POTENTIAL.  
 SQ SEQUENCE 506 AA; 57438 MW; 39B765F319235AD1 CRC64;

Query Match 17.4%; Score 1020; DB 1; Length 506;  
 Best Local Similarity 47.9%; Pred. No. 1e-47;  
 Matches 212; Conservative 76; Mismatches 149; Indels 6; Gaps 4;

QY 29 FKKILVANGETAVRAAFRAALTEGAATVAIYPREDRGSFHRSFASAVRIGTGGSPVKAY 88  
 DB 2 FSKILVANGETAVRVNRRACRLGKTGVYSSADKRAFRHYADCCYIG-KADPRDSY 60  
 QY 89 LDIDEITGAARKVKADAIYPGYGFLSENALARECAENGITFIGTPTEVDLDTGKSRV 148  
 DB 61 LNIIDRIEIVAKKSGAEAIHPGYGFLAENAEFAECRCCEEGVFIGPSPEVIRIAGSKVRSR 120



DE carboxylase (EC 6.4.1.2)) (ACC).  
 GN ACCC OR HI0972  
 OS Haemophilus influenzae.  
 CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 CC Haemophilus.  
 CC NCBI\_TaxID=727;  
 RX  
 RN  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Rd / KW20 / ATCC 51907;  
 RA MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RA rd.";  
 RL Science 269:496-512(1995).  
 CC  
 CC -1- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A  
 CC CARBOXYLASE COMPLEX; FIRST, BIOTIN CARBOXYLASE CATALYZES THE  
 CC CARBOXYLATION OF THE CARRIER PROTEIN AND THEN THE TRANSCARBOXYLASE  
 CC TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + biotin-carboxyl-carrier protein + CO(2)  
 CC = ADP + phosphate + carboxybiotin-carboxyl-carrier protein.  
 CC -1- PATHWAY: Long-chain fatty acid biosynthesis; first step.  
 CC -1- SUBUNIT: ACETYL-COA CARBOXYLASE IS AN HETERODIMER OF BIOTIN  
 CC CARBOXYL CARRIER PROTEIN, BIOTIN CARBOXYLASE AND THE TWO SUBUNITS  
 CC OF CARBOXYL TRANSFERASE IN A 2:2 COMPLEX (BY SIMILARITY).  
 CC -1- SIMILARITY: TO OTHER BIOTIN-DEPENDENT ENZYMES AND CARBAAMOYL-  
 CC PHOSPHATE SYNTHETASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; U32778; AAC22632.1; -  
 DR HSSP; P24182; 1BNC.  
 DR TIGR; HI0972; -  
 DR InterPro; IPR004549; AccC.  
 DR Pfam; PF00289; CPSase\_Lchain; 1.  
 DR Pfam; PF02785; Biotin\_carb\_C; 1.  
 DR Pfam; PF02786; CPSase\_L\_D2; 1.  
 DR TIGRFAMs; TIGR00514; ACCC; 1.  
 DR PROSITE; PS00866; CPSASE\_1; 1.  
 DR PROSITE; PS00867; CPSASE\_2; 1.  
 KW Fatty acid biosynthesis; Ligase; Biotin; ATP-binding;  
 KW Complete proteome. 168  
 FT NP\_BIND 163 168 ATP (POTENTIAL).  
 FT ACT\_SITE 292 292 POTENTIAL.  
 FT SEQUENCE 448 AA; 49108 MW; 28497E2A31ED956D1 CRC64;  
 Query Match 16.2%; Score 950; DB 1; Length 448;  
 Best Local Similarity 48.3%; Pred. No. 4.9e-44;  
 Matches 212; Conservative 57; Mismatches 152; Indels 18; Gaps 9;  
 QY 30 KKLIVANGETAVRAAETGAATVATVYPRDRGSRFSRFASEAVRIGTEGSPKAYL 89  
 DB 3 EKVIANRGEIATRLRACKELGKTKTAVHSTADRLKHLVLADETICIGPAPS-AKSYL 61  
 QY 90 DIDEIIGAIAKVKADAIYGVGFLENALARECAENGTTFGTPTPEVLDLTGDKSRAVT 149  
 DB 62 NIPATIAAAEVTGDAIHPGVGFLENADFAEQVRSRGTFTGPTADVIRLMGDKVSAIK 121  
 QY 150 AAKKAGLPVLAEST-PSKNIDDIIVS---AEGQTYPIEVKAVAGGGGRGMRVSPDELRL 205

Db 122 AMKKAGVPCVPGSDGPSN--DIARKKEIAKRIGYPIIIKASGGGGGRGMRVVRSEDALE 179  
 QY 206 KLATEASREAAFGDGSVYVERAVINPOHIEVILGORTGEVHLHYERDCSLORRHOKV 265  
 Db 180 ESIAMTKAEAKAFAFNDWMVMEKYLENPRHVEIQVLADTHGNVYLAERDCSMQRHOKV 239  
 QY 266 VEIAPQHLDPDLRI---CADAVKFCRSIGYAGAGTVEFLVDEKGNHVFTEMNPRIOV 322  
 Db 240 VBEAPAGPTEVRRDGRGRCANA---CVEICRGAGTGEFLY-ENGEPYFIEMNTRIOV 295  
 QY 323 EHTVTEVEVDLVKAOMRLAAGATLKEGLTQDKIKTHGAALQCRITTEDPNNGRPRDT 382  
 Db 296 EHPVTEMITGVDLVKRQLRAAGL---PTSFKEEDIKVKGHAMECRINAEDPKT-FLPSP 351  
 QY 383 GTTATVRSRPGAGVRDGAAGLQGGTAHFDMSLVKMTCRGSDFFETAVARAQALAEFTV 442  
 Db 352 GKVNHLSRSGGLGVWRWDVHYGYTPVPHYSMIAKLITYGDTREVAIRRMQNALSETII 411  
 QY 443 SGVATNIGELFRALLREEDF 461  
 Db 412 DGIKTNIPHLHELLEDEF 430  
 RESULT 12  
 ACCC\_ECOLI  
 ID ACCC\_ECOLI STANDARD; PRT; 449 AA.  
 AC P24182.  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Biotin carboxylase (EC 6.3.4.14) (A subunit of acetyl-CoA  
 DE carboxylase (EC 6.4.1.2)) (ACC).  
 GN ACCC OR PABG OR B3256.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OC NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RC STRAIN-K12;  
 RX MEDLINE=92052166; PubMed=1682920;  
 RA Kondo H., Shiratsuchi K., Yoshimoto T., Masuda T., Kitazono A.,  
 RA Tsuru D., Anai M., Sekiguchi M., Tanabe T.;  
 RT "Acetyl-CoA carboxylase from Escherichia coli: gene organization and  
 RT nucleotide sequence of the biotin carboxylase subunit.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:9730-9733(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92112819; PubMed=1370469;  
 RA Li S.-J., Cronan J.E. Jr.;  
 RT "The gene encoding the biotin carboxylase subunit of Escherichia coli  
 RT acetyl-CoA carboxylase.";  
 RL J. Biol. Chem. 267:855-863(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Best E.A., Knauf V.C.;  
 RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [5]  
 RP SEQUENCE OF 1-12.  
 RC STRAIN=K12 / EMG2;  
 RX MEDLINE=97443975; PubMed=9298646;  
 RA Link A.J., Robison K., Church G.M.;

RT "Comparing the predicted and observed properties of proteins encoded  
 RL in the genome of *Escherichia coli* K-12.";  
 RN Electrophoresis 18:1259-1313(1997).  
 RX X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).  
 RA MEDLINE=94347758; PubMed=7915138;  
 RA Waldrop G.L., Rayment I., Holden H.M.;  
 RT "Three-dimensional structure of the biotin carboxylase subunit of  
 RL acetyl-CoA carboxylase.";  
 RN Biochemistry 33:10249-10256(1994).  
 RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).  
 RA MEDLINE=20283656; PubMed=10821865;  
 RA Thoden J.B., Blanchard C.Z., Holden H.M., Waldrop G.L.;  
 RT "Movement of the biotin carboxylase B-domain as a result of ATP  
 binding.";  
 RN J. Biol. Chem. 275:16183-16190(2000).  
 CC -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A  
 CC CARBOXYLASE COMPLEX; FIRST, BIOTIN CARBOXYLASE CATALYZES THE  
 CC CARBOXYLATION OF THE CARRIER PROTEIN AND THEN THE TRANSCARBOXYLASE  
 CC TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA.  
 CC -!- CATALYTIC ACTIVITY: ATP + biotin-carboxyl-carrier protein + CO(2)  
 CC = ADP + phosphate + carboxybiotin-carboxyl-carrier protein.  
 CC -!- PATHWAY: Long-chain fatty acid biosynthesis; first step.  
 CC -!- SUBUNIT: ACETYL-COA CARBOXYLASE IS AN HETEROHEXAMER OF BIOTIN  
 CC CARBOXYL CARRIER PROTEIN, BIOTIN CARBOXYLASE AND THE TWO SUBUNITS  
 CC OF CARBOXYL TRANSFERASE IN A 2:2 COMPLEX.  
 CC -!- SIMILARITY: TO OTHER BIOTIN-DEPENDENT ENZYMES AND CARBAMOYL-  
 CC PHOSPHATE SYNTHETASES.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; M79446; AAA23748.1; -.  
 DR EMBL; M80458; AAA23409.1; -.  
 DR EMBL; M83198; AAA23746.1; -.  
 DR EMBL; U18997; AAA58059.1; -.  
 DR EMBL; AE000404; AAC76288.1; -.  
 DR PIR; JS0632; JS0632.  
 DR PIR; JS0687; JS0687.  
 DR PDB; 1BNC; 30-AUG-95.  
 DR PDB; 1DV1; 09-JUN-00.  
 DR PDB; 1DV2; 09-JUN-00.  
 DR Ecogene; EG10276; accC.  
 DR InterPro; IPR004549; AccC.  
 DR InterPro; IPR000901; CPSase.  
 DR Pfam; PF00289; CPSase\_L\_chain; 1.  
 DR Pfam; PF02785; Biotin\_carb\_C; 1.  
 DR Pfam; PF02786; CPSase\_L\_D2; 1.  
 DR TIGRFAMS; TIGR00514; accC; 1.  
 DR PROSITE; PS00866; CPSASE.1; 1.  
 DR PROSITE; PS00867; CPSASE.2; 1.  
 KW Fatty acid biosynthesis; Ligase; Biotin; ATP-binding; 3D-structure;  
 KW Complete proteome.  
 FT NP\_BIND 163 168 ATP (POTENTIAL).  
 FT ACT\_SITE 292 292 POTENTIAL.  
 FT CONFLICT 260 261 CA -> SR (IN REF. 2).  
 FT CONFLICT 313 313 L -> M (IN REF. 1).  
 SQ SEQUENCE 449 AA; 49320 MW; 68C55F10ACB4F170 CRC64;

Query Match 16.0%; Score 940; DB 1; Length 449;  
 Best Local Similarity 46.8%; Pred. NO. 1.7e-43;  
 Matches 203; Conservative 66; Mismatches 155; Indels 10; Gaps 6;

OY 31 KILVARGEIAVRAFAALTECAATVAIYPREDSCFSFRHSFASEAVRICTEGSPVKAYLD 90  
 DB 4 KIVIANRGEIALRIILRACKELGIKTVAHVSSADRLKHLVLADETVCIGPAPS-VKSYLN 62

OY 91 IDEIIAAGKVKADAIYPGYGFLSENAQLARECAENGITFIGPTPEVLDLTGDKSAVTA 150  
 DB 63 IFAITSAAEITGVAIHPCYGFSENAFAEQVERSGFIFIGPKAETIRLMGDKVSAIAA 122  
 OY 151 AKKAGLPVLAESTPSKNIDDIKVS---AEGQTVPIFVAVAGGGGRMRFPSSPDELKRL 207  
 DB 123 MKKAGVPCVPDGLG-DMDKNRAIKRIGYPIVIAKASGGGGGRMRVVRGDAELAQ 181  
 OY 208 ATEASREAAAFDCGYYVERAVINPOHIEVOILGDRTEGVHLYERDCSLORRHOKVVE 267  
 DB 182 ISMTRAEAKAFAFNDVMYMEKYLENPRHVEIQVLADQGNIAIYLAERDCSMORRHOKVVE 241  
 OY 268 IAPAQLDPELRDRICADAVKFCRSIGYOGAGTVEELVDEKGNHVTIENNPRIQVHTVT 327  
 DB 242 EAPAGITELRLRYIGERCACAKVDIGYRGAGTFEFLF-ENGFEYFIEMNTRIQVEHPVT 300  
 OY 328 EEVTEVDLVKAQMLAAGATLKLGLTKTHGAALOCRTTPEDNNGFRPDGTGTTA 387  
 DB 301 EMITGVDLIKEQLRIAG--QPLSIKQBEVHVGRHAVECRINAEDPNT-FLPSPGKITR 356  
 OY 388 YRSPGCGAVRLDGAALGGEITAHFDSMLVKTCTGSDPETAFAVARAQLAEFTVSGVAT 447  
 DB 357 FHAPGGGVRWESHYGYTVPPYDYMIGKLCICYGENRDVAIARMKNALQELIIDGIKT 416  
 OY 448 NIGFLRALLREEDF 461  
 DB 417 NVDLQIRINWENF 430

## RESULT 13

ACCC\_ECO57 STANDARD; PRT; 449 AA.  
 AC O8X9B6;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Biotin carboxylase (EC 6.3.4.14) (A subunit of acetyl-CoA  
 DE carboxylase (EC 6.4.1.2)) (ACC).  
 GN ACCC OR 24616 OR ECS4128.  
 OS *Escherichia coli* O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC *Escherichia*.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 CC -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE ACETYL COENZYME A  
 CC CARBOXYLASE COMPLEX; FIRST, BIOTIN CARBOXYLASE CATALYZES THE  
 CC CARBOXYLATION OF THE CARRIER PROTEIN AND THEN THE TRANSCARBOXYLASE  
 CC TRANSFERS THE CARBOXYL GROUP TO FORM MALONYL-COA.  
 CC -!- CATALYTIC ACTIVITY: ATP + biotin-carboxyl-carrier protein + CO(2)  
 CC = ADP + phosphate + carboxybiotin-carboxyl-carrier protein.  
 CC -!- PATHWAY: Long-chain fatty acid biosynthesis; first step.



QY 30 KKILVANGEIAVRAALETGAATVAIYPREDGSRHRSFASAVRIGTEGSPVKAYL 89  
 Db 3 EKVLIANRGEIALRLRKLKELGKIKTAVHSTADRELMLHLSLADESVICG-PAPATQSYL 61  
 QY 90 DIDEIIGAIAKKVADAIYPGVGFLENQAQARECAENIGTIFGPTPEVLDLTGDKSRVAVT 149  
 Db 62 QIPALIAAAEVTVGATAIIPGVGFLENQAQARECAENIGTIFGPTPEVLDLTGDKSRVAVT 121  
 QY 150 AAKKAGLPVL--AESTPSKNTDDIVKSAEGOTYTFVFAVAGGGGRGMRFFSSPDRLKL 207  
 Db 122 AMKRAVTVPGSDGLPEDEETALAIAREVGYPIIIKAGGGGRGMRFFSSPDRLKL 181  
 QY 208 ATEASREAEAFGDSVYVVERAVINPQHIIEVOILGDRTEGVVHLVYERDCSLQRHKKVVE 267  
 Db 182 AKLTRTEAGAAFGNPMVYLEKFLTPRHEVVEQVLSGOGNAIHLGDRDCSLQRHKKVIE 241  
 QY 268 IAPQHLPELRDRICADAVFCRSIGYOGAGTVEFLVDEKGNHVFIEPNRIQVEHVT 327  
 Db 242 EAPAGIDEKARQEVFAFCVQACIEIGYRGAGTFFELY-ENGRFFYFIEMNTRVQVHEPVS 300  
 QY 328 EEVTEVDLVKQMRLLAAGATLKLGLTQDKIKTHGAALQCRITTEDPNNGFRPDTGTITA 387  
 Db 301 EMWTVGVIVKEMLRATSG---EKLISROEDVVIRGHAELECRINAEDPKT-TWMPSPGKVKH 356  
 QY 388 YRSPGAGVRIDGAQALGGEITAHFDSMLVKMTCRGSDFETAVARAQALAEFTVSGVAT 447  
 Db 357 FHAPGGNGVRVDSHLYSGVSPNNYDLSGVKITVYTGADRDALARMNALDELIVDGIKT 416  
 QY 448 NIGFLRALLREDF 461  
 Db 417 NTELHKDLVRDAAF 430

RESULT 15  
 MCCA\_HUMAN STANDARD; PRT; 725 AA.  
 ID MCCA\_HUMAN Q9RQ3; Q9H959; Q9NS97;  
 AC 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Methylcrotonyl-CoA carboxylase alpha chain, mitochondrial precursor  
 DE (EC 6.4.1.4) (3-Methylcrotonyl-CoA carboxylase 1) (MCCase alpha  
 subunit) (3-methylcrotonyl-CoA:carbon dioxide ligase alpha subunit).  
 GN MCC1 OR MCCA.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND VARIANTS MCGI ARG-325 AND SER-385.  
 RX PubMed=11170888;  
 RA Gallardo M.E., Desviat L.R., Rodriguez J.M., Esparza-Gordillo J.,  
 Perez-Cerdas C., Perez B., Rodriguez-Pombo P., Criado O., Sanz R.,  
 Morton D.H., Gibson K.M., Le T.P., Ribes A., Rodriguez de Cordoba S.,  
 Ugarte M., Penalba M.A.;  
 RT "The molecular basis of 3-methylcrotonylglycinuria, a disorder of  
 RT leucine catabolism".  
 RL Am. J. Hum. Genet. 68:334-346(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND VARIANT HIS-464.  
 RX MEDLINE=21295033; PubMed=11401427;  
 RA Obata K., Fukuda T., Morishita R., Abe S., Asakawa S., Yamaguchi S.,  
 Yoshino M., Ihara K., Murayama K., Shigemoto K., Shimizu N., Kondo I.;  
 RT "Human biotin-containing subunit of 3-methylcrotonyl-CoA carboxylase  
 RT gene (MCCA): cDNA sequence, genomic organization, localization to  
 RT chromosomal band 3q27, and expression".  
 RL Genomics 72:145-152(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A., AND VARIANT MCGI PHE-535.  
 RX MEDLINE=21299419; PubMed=11405611;  
 RA Holzinger A., Roeschinger W., Lagler F., Mayerhofer P.U., Lichtner P.,  
 RA Kattenfeld T., Thuy L.P., Nyhan W.L., Koch H.G., Muntau A.C.,  
 RA Roscher A.A.;

"Cloning of the human MCCA and MCB genes and mutations therein reveal  
 the molecular cause of 3-methylcrotonyl-CoA: carboxylase  
 Hum. Mol. Genet. 10:1299-1306(2001).  
 [4]  
 RP SEQUENCE FROM N.A., AND VARIANTS MCGI VAL-289; SER-385; PRO-437 AND  
 RP HIS-532.  
 RX PubMed=11181649;  
 RA Baumgartner M.R., Almashanu S., Suormala T., Obie C., Cole R.N.,  
 Packman S., Baumgartner E.R., Valle D.;  
 RT "The molecular basis of human 3-methylcrotonyl-CoA carboxylase  
 RT deficiency".  
 RL J. Clin. Invest. 107:495-504(2001).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzûki Y.,  
 Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,  
 Matsumura Y., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,  
 Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;  
 RT "NEDO human cDNA sequencing project".  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skeletal muscle;  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: ATP + 3-methylcrotonyl-CoA + HCO(3)(-) - ADP +  
 CC phosphate + 3-methylglutaconyl-CoA.  
 CC -1- COFACTOR: Biotin.  
 CC -1- PATHWAY: Leucine catabolism.  
 CC -1- SUBUNIT: Probably a dodecamer composed of six biotin-containing  
 CC alpha subunits and six beta subunits.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.  
 CC -1- DISEASE: Defects in MCC1 are the cause of 3-  
 CC methylcrotonylglycinuria type I (MCGI, CGA or CG2). MCGI is a  
 CC recessive disease that is characterized by muscular hypotonia and  
 CC atrophy, probably of spinal origin.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF310972; AAG53095.1; -  
 CC EMBL; AB029826; BAA99407.1; -  
 CC EMBL; AF297332; AAK67986.1; -  
 CC EMBL; AF310339; AAG50245.1; -  
 CC EMBL; AK023051; BAB14377.1; -  
 CC EMBL; BC004214; AAH04214.1; -  
 CC EMBL; BC004187; AAH04187.1; -  
 CC Genbank; HGNC:6936; MCC1.  
 CC MIM; 210200; -  
 CC InterPro; IPR001882; Biotin\_attach.  
 CC InterPro; IPR000089; Biotin\_lipoyl.  
 CC InterPro; IPR000901; CPSase.  
 CC Pfam; PF00289; CPSase\_L\_chain; 1.  
 CC Pfam; PF00364; biotin\_lipoyl; 1.  
 CC Pfam; PF02785; biotin\_carb\_C; 1.  
 CC Pfam; PF02786; CPSase\_L\_D2; 1.  
 CC PROSITE; PS00188; BIOTIN; 1.  
 CC PROSITE; PS00867; CPSASE\_2; 1.  
 CC Mitochondrion; Ligase; Biotin; ATP-binding; Transit peptide;  
 KW Disease mutation; Polymorphism.  
 FT TRANSIT 1 47  
 FT CHAIN 48 725 MITOCHONDRION (POTENTIAL).  
 FT NP\_BIND 209 214 METHYLCROTONYL-COA CARBOXYLASE ALPHA  
 FT ACT\_SITE 339 339 CHAIN.  
 FT BINDING 681 681 ATP (POTENTIAL).  
 FT DOMAIN 538 541 BIOTIN (BY SIMILARITY).  
 FT POLY-SER. 541





GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 26, 2003, 18:57:30 ; Search time 110 seconds

(without alignments)  
2167.241 Million cell updates/sec

Title: US-09-974-973-2

Perfect score: 5865

Sequence: 1 MTAITLGLGLLKGIITLVST.....RVVVPATKVEGDLIVVVS 1157

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5759	98.2	1140	2	O54587
2	5297.5	90.3	1139	2	O8RQL2
3	3797	64.7	1124	16	O9RK64
4	3786.5	64.6	1127	2	O9F843
5	3713.5	63.3	1127	16	P95127
6	3484	59.4	1124	2	O50450
7	2535.5	43.2	1150	16	O9K9M0
8	2530	43.1	1148	16	O9KWD4
9	2523.5	43.0	1180	13	O9DDT1
10	2509	42.8	1147	2	P94448
11	2507.5	42.8	1192	3	O9HES8
12	2494	42.5	1146	16	O9Y846
13	2491	42.5	1146	16	O9ZCW1
14	2490.5	42.5	1144	16	O97FR7
15	2489	42.4	1150	16	O99UY8
16	2485.5	42.4	1193	3	O93918

17	2474.5	42.2	1174	16	Q8UBX3
18	2473.5	42.2	1152	16	O98F27
19	2465.5	42.0	1158	16	O8YJ20
20	2461	42.0	1175	5	O17732
21	2447	41.7	1185	3	O90UE1
22	2441	41.6	1181	5	O9XZ00
23	2439	41.6	1185	3	P78822
24	2435.5	41.5	1175	3	O8X1T3
25	2435	41.5	1152	16	O92L13
26	2433.5	41.5	1196	5	O9I7E9
27	2417.5	41.2	1195	5	O16921
28	2413	41.1	1137	16	O9CHQ7
29	2405	41.0	1137	2	O9RAT6
30	2401	40.9	1154	2	O59740
31	2119.5	36.1	984	2	O9XBJ1
32	2021	34.5	935	11	O62043
33	1933.5	33.0	920	2	O9RW05
34	1071.5	18.3	477	16	O67449
35	1040	17.7	472	16	O67483
36	1005.5	17.1	453	2	O54755
37	982.5	16.8	384	2	O07640
38	979.5	16.7	471	16	O9HTD0
39	978.5	16.7	448	16	O55160
40	973.5	16.6	1095	16	O9I3U4
41	970.5	16.5	1338	5	O9GQ08
42	969.5	16.5	493	17	O8TSX0
43	966.5	16.5	493	1	O9HH17
44	966	16.5	573	2	O9FAF5
45	963.5	16.4	1078	16	Q9A3J0

#### ALIGNMENTS

RESULT 1

O54587 ID O54587 PRELIMINARY; PRT; 1140 AA.  
AC O54587;  
DT 01-JUN-1998 (Tremblrel. 06, Created)  
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Pyruvate carboxylase.  
GN PYC.  
OS Corynebacterium glutamicum (Brevibacterium flavum).  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
OC Corynebacterium.  
OX NCBI\_TaxID=1718;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=21253;  
RA Koffas M.A.G., Ramamoorthi R., Pine W.A., Sinskey A.J.,  
RA Stephanopoulos G.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13032;  
RA Peters-Wendisch P.G., Kreutzer C., Kalinowski J., Patek M., Sahn H.,  
RA Eikmanns B.J.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- COFACTOR: BIOTIN (BY SIMILARITY).  
DR EMBL; AF038548; AAB92588.1; -  
DR EMBL; Y09548; CAA70739.1; -  
DR HSSP; P24182; IBNC.  
DR InterPro; IPR001882; Biotin\_attach.  
DR InterPro; IPR000089; Biotin\_lipeyl.  
DR InterPro; IPR000901; CPase.  
DR InterPro; IPR000891; HMGL-like.  
DR InterPro; IPR003379; PYC\_OAD.  
DR Pfam; PF02785; Biotin\_carb\_C; 1.  
DR Pfam; PF00364; biotin\_lipeyl; 1.  
DR Pfam; PF00289; CPase\_L\_chain; 1.  
DR Pfam; PF02786; CPase\_L\_D2; 1.

DR	Pfam: PF00682; HMGL-like; 1.	DR	Pfam: PF02436; PYC_OADA; 1.	DR	TIGRFAMS: TIGR01235; Pyruv_carbox; 1.	DR	TIGRFAMS: TIGR01235; Pyruv_carbox; 1.	DR	PROSITE; PS00188; BPTIN; 1.	DR	PROSITE; PS00188; BPTIN; 1.	DR	PROSITE; PS00867; CTSASE_2; UNKNOWN_1.	KW	Biotin: Ligase; Pyruvate.	KW	Biotin: Ligase; Pyruvate.	SQ	SEQUENCE 1140 AA; 123102 MW; 2A6D4B4ED2FEB531 CRC64;	Query Match	98.2%; Score 5759; DB 2; Length 1140;	Best Local Similarity	99.4%; Pred. No. 4.6e-300;	Matches 1133; Conservative	5; Mismatches 2; Indels 0; Gaps																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
Qy	18	VSTTSSTLPAFKKILVANRGEIAVRAFRAL	ETGAATVAIYPREDRGSFHRSFASEAVR	77	1	MSTTSSTLPAFKKILVANRGEIAVRAFRAL	ETGAATVAIYPREDRGSFHRSFASEAVR	60	78	IGTEGSPVAKLDI	ETIIGA	KKADAI	YPGVGFLSEN	QAQARECA	ENGITFI	GI	PTPEV	137	61	IGTEGSPVAKLDI	ETIIGA	KKADAI	YPGVGFLSEN	QAQARECA	ENGITFI	GI	PTPEV	120	138	LDLTGDKSR	AVTA	AKKAGLP	VA	LES	PSK	NI	DDI	V	KS	AG	OT	Y	P	I	V	K	A	V	G	G	G	R	M	F	197	121	LDLTGDKSR	AVTA	AKKAGLP	VA	LES	PSK	NI	DDI	V	KS	AG	OT	Y	P	I	V	K	A	V	G	G	G	R	M	F	180	198	VSSPDEL	R	K	L	A	T	A	S	R	E	A	E	A	F	G	D	S	V	Y	R	A	V	I	N	P	O	H	I	E	V	O	I	L	G	D	T	E	V	H	L	Y	E	R	D	C	S	257	181	VASPDEL	R	K	L	A	T	A	S	R	E	A	E	A	F	G	D	S	V	Y	R	A	V	I	N	P	O	H	I	E	V	O	I	L	G	D	T	E	V	H	L	Y	E	R	D	C	S	240	258	LORRHO	K	V	E	I	A	P	A	Q	H	L	D	P	E	L	R	I	C	A	D	A	V	K	C	R	S	I	G	Y	Q	A	G	T	V	E	F	L	V	D	E	K	N	H	V	E	I	E	M	N	317	241	LORRHO	K	V	E	I	A	P	A	Q	H	L	D	P	E	L	R	I	C	A	D	A	V	K	C	R	S	I	G	Y	Q	A	G	T	V	E	F	L	V	D	E	K	N	H	V	E	I	E	M	N	300	318	PRIOVHT	V	T	E	E	V	D	L	V	K	A	Q	M	R	L	A	A	G	A	T	K	E	L	G	I	T	Q	D	K	I	K	H	G	A	L	O	C	R	I	T	E	D	P	N	N	G	377	301	PRIOVHT	V	T	E	E	V	D	L	V	K	A	Q	M	R	L	A	A	G	A	T	K	E	L	G	I	T	Q	D	K	I	K	H	G	A	L	O	C	R	I	T	E	D	P	N	N	G	360	378	FRPD	T	G	T	I	T	A	R	S	P	G	G	A	G	V	R	L	D	G	A	A	Q	L	G	G	E	I	T	A	H	F	S	D	S	M	L	V	K	M	T	C	R	G	S	D	E	F	A	V	A	R	A	Q	R	A	437	361	FRPD	T	G	T	I	T	A	R	S	P	G	G	A	G	V	R	L	D	G	A	A	Q	L	G	G	E	I	T	A	H	F	S	D	S	M	L	V	K	M	T	C	R	G	S	D	E	F	A	V	A	R	A	Q	R	A	420	438	A	E	F	T	S	G	V	A	T	N	I	G	F	R	A	L	L	R	E	E	D	F	T	S	K	R	I	A	T	G	I	G	B	H	P	H	L	L	A	P	P	A	D	D	P	Q	G	R	I	L	D	Y	497	421	A	E	F	T	S	G	V	A	T	N	I	G	F	R	A	L	L	R	E	E	D	F	T	S	K	R	I	A	T	G	I	G	B	H	P	H	L	L	A	P	P	A	D	D	P	Q	G	R	I	L	D	Y	480	498	T	N	K	P	H	C	V	R	P	K	D	V	A	A	P	I	D	K	L	P	N	I	K	D	L	P	R	G	S	R	D	R	L	Q	L	G	P	A	A	F	A	R	L	R	E	Q	A	L	A	V	T	D	T	557	481	T	N	K	P	H	C	V	R	P	K	D	V	A	A	P	I	D	K	L</

Db	901	AADPÖKYDII	PDSVIAFURJGELGNPPGGWPEURTRALCEKRGSEKAPLTFEVPESQAHILDA	960			
Qy	978	DOSKERNSNLRLLFPKPTEEFLEHRRRRCNGTSSALDDREFFYGLVGEGRITLRLPDVRTP	1037				
Db	961	DOSKERNSNLRLLFPKPTEEFLEHRRRRCNGTSSALDDREFFYGLVGEGRITLRLPDVRTP	1020				
Qy	1038	LLVLRLDAI	SEPDDOKGRNVVNANGOTRPMRVDRSRVSVESTATAEKADSSNKHGVAAPFA	1097			
Db	1021	LLVLRLDAI	SEPDDOKGRNVVNANGOI RPMRVDRSRVSVESTATAEKADSSNKHGVAAPFA	1080			
Qy	1098	GVTVTVAEGDEVKAGDAVAI	IEAMKEATITASVDGKTIRVVVPAATKVSGDDLIVVVS	1157			
Db	1081	GVTVTVAEGDEVKAGDAVAI	IEAMKEATITASVDGKTIDRVVVPAATKVSGDDLIVVVS	1140			
<hr/>							
RESULT 2							
<hr/>							
ID	QBOL2	PRELIMINARY; PRT: 1139 AA.					
AC	QBOL2;						
DT	01-JUN-2002 (T-EMBLrel. 21, Created)						
DT	01-JUN-2002 (T-EMBLrel. 21, Last sequence update)						
DE	01-JUN-2002 (T-EMBLrel. 21, Last annotation update)						
DI	pyruvate carboxylase.						
GN	PC.						
OS	Corynebacterium efficiens.						
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;						
OC	Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;						
OC	Corynebacterium.						
OX	NCBI_TaxID=152794;						
RN	[1]						
RP	SEQUENCE FROM N.A.						
RA	Akiyoshi N., Nonaka G., Kimura E., Kawahara Y., Sugimoto S.;						
RT	"Corynebacterium efficiens pyruvate carboxylase (pyc) gene, complete						
RT	CDS."						
RL	Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.						
DR	EMBL; AB083299; BAB88903.1;						
SQ	SEQUENCE 1139 AA; 123068 MW; BA7023134519FAAA CRC64;						
<hr/>							
Query Match 90.3%; Score 5297.5; DB 2; Length 1139;							
Best Local Similarity 90.9%; Pred. No. 2.5e-275;							
Matches 1035; Conservative 44; Mismatches 58; Indels 1; Gaps							
Qy	20	THTSSTLPAPFKKILVANRGEIAVRAFRALETGAATVAIYPREDRSPHRSFASEAVRIG	79				
Db	3	TTPTSTLPAPFKKILVANRGEIAVRAFRAYETGAATVAIYPREDRSPHRSFASEAVRIG	62				
Qy	80	TEGSPVKAYLDIDEIIGAAKYKADAIYPGYGLFSENQAQLARECAENGITFTGPTPEVLD	139				
Db	63	TEGSPVKAYLDIDEINNAKKYKADAVIPGYGLFSENQAQLARECAENGITFTGPTPEVLD	122				
Qy	140	LTGDKSRVTAANKAGLPVLAESTPSKNTDDIVKSABEGTYPIFVKAVAGGSGRGMRFVS	199				
Db	123	LTGDKSKAVSAANKAGLPVLAESTPSTDIDEIVKSAEGTYPIFVKAVAGGSGRGMRFVE	182				
Qy	200	SPDELRLKATASREAFAFGDGSVYVERAVINPOHIEVQILGDRTEGVHLHYERDCSLQ	259				
Db	183	KPEDRLRELAREAREAFGDGSVYVERAVIKPHIEVQILGHDTGVDVHLYERDCSLQ	242				
Qy	260	RRHKQVVEIPAQHLDPFLDRICADAVKFCRSICYQGAGTVFEFLVDEKGHNHFVIENMR	319				
Db	243	RRHKQVVEIPAQHLDPFLDRICADAVKFCRSICYQGAGTVFEFLVDEAGNHVFVIENMR	302				
Qy	320	IQVEHTVTEEVTVSDLVKACMRLAAGATLKEGLTGQDKIKTHGAALQCRIITPEDPNNGFR	379				
Db	303	IQVEHTVTEEVTVSDLVKACQMHLAAGATLKEGLTGQDKITTHGAALQCRIITPEDPSNNFR	362				
Qy	380	PDTGTITAYRSPGAGVRLDGAQAQLGGETAITAFDMSLVKMTCRGSDFTAVARAQRALAE	439				
Db	363	PDTGVITAYRSPGAGVRLDGAQAQLGGETAITAFDMSLVKMTCRGSDFTAVSRAQRALAE	422				
Qy	440	FTVSGVATNI GFURLLRBEDFTSKRIATGFTGDHPHLLQAAPPADDEOGRILDYADVTV	499				

```
Db 423 FNVSGVATNIGFRLALLREEDFTKRRIDGTGSHQHLLQAPPADDEQGRILEYLAADVTV 482
QY 500 NKPHGVPRKDVAAPIDKLNKIDPLPRGSRDRKLGKGPAAFAARDLREODALAVDTTFR 559
Db 483 NKPHGERP-ETARPIEKLEVENIPLPRGSRDRKLGKGPFGFARDLREODALAVDTTFR 541
QY 560 DAHQSLLATRVRSFALKPAAEAVAKLTPELLSVSEANGGATYDVAMRFLFEDPWRDLDEL 619
Db 542 DAHQSLLATRVRSFALTPAARAVAKLTPELLSVSEANGGATYDVAMRFLFEDPWRDLDEL 601
QY 620 EAMPNNIOMLGRNTVGYTYPDSVCRAFYKEAAASSGVDFRIFDALNDVDSQMPAID 679
Db 602 EAMPNNIOMLGRNTVGYTYPDSVCRAFYQEAASGVDFRIFDALNDVDSQMPAID 661
QY 680 AVLETNTAEEVAMAYSGDLSNPNEKLYLDYVYKMAEEIVKSGAHILAIAKOWAGLLRPA 739
Db 662 AVLETCTVAEAMAYSGDLSNPGEKLYLDYVYLNLAQIVDSGAHILAIAKOWAGLLRRA 721
QY 740 AVTKLVTLRRRFDLPVHVHTHTAGGQATYFAAAQAGADAVDGSAPLSGTTSPQSL 799
Db 722 AAPKLVTLRRRFDLPVHVHTHTAGGQATYLAANAAGADAVDAASAPLSGTTSPQSMS 781
QY 800 AIVAAFAHTRRDTGLSLEAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGGQLSN 859
Db 782 ALVAAFAHTRRDTGULQAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGGQLSN 841
QY 860 LRAQATAGLADRFEIENYAAVNEMLGRPTKVTSSKVVGDALHLVYAGVDPADFAA 919
Db 842 LRAQAVAGLADRFEIENYAAVNEMLGRPTKVTSSKVVGDALHLVYAGVSPEDFAA 901
QY 920 DPKYDIPDSVIAFLRGELGNPPGWPPELRTALREGRSEKAPLTEVPEEQAHLDADD 979
Db 902 DPKYDIPDSVIAFLRGELGTPGGWPPELRTALREGRSOGKAPLAEIPAEQAHLDSD 961
QY 980 SKERNSLNRLFLFPKPTFEFLHRRRFGNTSALDRFFYGLVEGRETLIRLPDVRTPLL 1039
Db 962 SAERGTNLNRLFLFPKPTFEFLHRRRFGNTSALDRFFYGLKEGREELIRLTGVSTPMV 1021
QY 1040 VRLDAISEDDKGMNVANVANGQIRPMRVDRSVESVTATKADSSNKGHVAAPFAGV 1099
Db 1022 VRLDAVSEDDKGMNVVNVANGQIRPKVDRSVESVTATKADATNKGHVAAPFAGV 1081
QY 1100 VTTVAEGDEVGADAVAIIEAMKMEATITASVDGKIERYVVPAAFKVEGGDLIVVVS 1157
Db 1082 VTTVVAEGDEIKAGDAVAIIIEAMKMEATITAPVDGVIDRVVVPAAFKVEGGDLIVVVS 1139

RESULT 3
Q9RK64
ID Q9RK64 PRELIMINARY; PRT; 1124 AA.
AC Q9RK64;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pyruvate carboxylase.
GN SC00546 OR ScF11.26C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_taxid=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
```

```
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL: AL132662; CAB59603.1; -.
DR HSP; P24182; IBCN.
DR InterPro; IPR000089; Biotin_lipoyl.
DR InterPro; IPR000901; CPSase.
DR InterPro; IPR000891; HMGL-like.
DR InterPro; IPR003379; PYC-OADA.
DR Pfam; PF02785; Biotin_carb_C; 1.
DR Pfam; PF00364; biotin_lipoyl; 1.
DR Pfam; PF00289; CPSase_L_D2; 1.
DR Pfam; PF02786; CPSase_L_D2; 1.
DR Pfam; PF00682; HMGL-like; 1.
DR Pfam; PF02436; PYC-OADA; 1.
DR TIGRFAMs; TIGR01235; pyruv_carbox; 1.
DR PROSITE; PS00867; CPSASE_2; UNKNOWN_1.
SQ SEQUENCE 1124 AA; 121071 MW; 1FEF4C4643954D31 CRC64;

Query Match 64.7%; Score 3797; DB 16; Length 1124;
Best Local Similarity 65.2%; Pred. No 6e-195;
Matches 735; Conservative 156; Mismatches 227; Indels 10; Gaps 6;

QY 29 FKKILVANGEIAVRAFAALETGAATVAIYPREDSGFHRSPASEAVRITGEGSPVKAY 88
Db 2 FRKVLVANRGEIAIRAIRAGYELGARTVAVFPEDRNSLRHLKADAEYIEGEGHPVAY 61
QY 89 LDIDEILGAKKAKDAIYPGYGLSENAQLARECEANGITFTGPTPEVLDTGDKSRAY 148
Db 62 LSVEEIVRAARRAGADAVYPGYGLSENPELARACEEAGITFVGSPARILELTGNKARAY 121
QY 149 TAAKAGLPVLAESTPSKNIDDIIVKSABGQTYPIFKAVAGGGGRGMRFPDDELRLKA 208
Db 122 AAAREAGVPVLGSSAPSTDVDELVRAADVDGFPFVKAVAGGGGRGMRVPEPAQLREAI 181
QY 209 TEASREAAAFGDSGVYVRAVINPQHIEVQILGDRTGVEVHLYERDCSLQRHQKVEI 268
Db 182 EAASREAAAFGDSVTVELEKAVVEPRHIEVQILADGEGDVIHLFERDCSVQRHQKVEL 241
QY 269 APAQHLDPELRDRICADAVKFCRSICVAGAGTVFELVDEKGNHVFTEMNPRIOVEHTVTE 328
Db 242 APAPNLDPALRERICADAVNFARQIGYRNAGTVFELVDRDGNHVFTEMNPRIOVEHTVTE 301
QY 329 EVTEVDLVRAQMRLAAGATLKEGLTQDKTKTHGAALQCRITTTEDPNNGFRPDTGTITAY 388
Db 302 EYTDVDLVQSOLRIAGQTLADLGLAQENITLRGAALQCRITTTEDPANGFRPDTGQISAY 361
QY 389 RSPGGAGVRLD-GAAQLGGEITAHFDSMLVKMTCRSDFFETAVARAQALAEFTSVGAT 447
Db 362 RSPGGGIRLDGGTTHAGTEISAHFDMSMLVKLCRGRDFTTAVNARRAFAEIRGAT 421
QY 448 NIGFLRALLREEDFTSKRTATGIGDHPHLLQAPPADDEQGRILDYLAADVTKNPKHGV 507
Db 422 NIPFLQAVLDDPDPFQAGRVTTSFIEQRPHLLTARHSADRGTKLITYLAADVTKNPKHGERP 481
```

```
Qy 508 KDVAAPIIDKLPNIKDLPLPGSRDRLLKQLGPAFAFARDLREDDALAVTDTTFRDAHQSLLA 567
Db 482 -ELVDPLTKLPTASAGEPAGSROLLAELGEGFARLRRESSTIGVDTTFRDAHQSLLA 540
Qy 568 TRVRSFALKPAAEAVAKLTPELLSVEAWGATYDVAMRFLFEDPDWDLDELREAMPNVI 627
Db 541 TRVTRKDLAVAPVAVARTLPLLSELCWGATYDVALLFLAEDPWERLALAREAVPNLCL 600
Qy 628 QMLLRGRNTVGYTPYDPSVCRFAFYKEAASSGVDFIRIFDALNDVDSQMRPAIDAVLETNTA 687
Db 601 QMLLRGRNTVGYTPYPTVEVTDFAFQEAATGIDIFRIFDALNDVEQMRPAIEAVROTGSA 660
Qy 688 VAEVAMAYSGDLSDPNEKLYTLDYLLKMAEIVKSGAHILAIKDMAGLLRPAAVTKLVTA 747
Db 661 VAEVALCYTADLSDPSERLYTLDYLLRLAQIVNAGAHVLAVKDMAGLLRPAAPATLVA 720
Qy 748 LRREFDLPVHVHTDTHAGGOLATYFAAAQAGADAVDCASAPLSGTTSSQPSLSAIVAAFDH 780
Db 721 LRREFDLPVHLHTDTHGGOLATYLAIAQAGADAVDCASAMAGTTSSQPSLSAIVAAFDH 780
Qy 808 TRRDTGLSLEAVSDLEPYWEAVRGLYLPFESGTPGPRGVYRHEIPCGOLSNLRACATL 867
Db 781 TERPTGLDQAVGLDLEPYWESVRKVYAPFEAGLASPTGRVYHHEIPCGOLSNLRCAVAL 840
Qy 868 GLADREFELIEDNYAAVNMGLRPTKVPSSKVVYCDLALHLVAGVDPADPAADPOKYDIP 927
Db 841 GLGRFEDIENYAAADRLMLRLVKTPTSSKVVYCDLALHLVAGVSPADPEQDPDRFIDIP 900
Qy 928 DSVTAFLRGLNCPGPGWPEPLTRALEGRSEKAPLTVPEVEEQAHLDADDKERNLSL 987
Db 901 DSVVGLFELGELGTTPGPGWPEPLRSKALGRAEAR-PLAELSEDDROGL----GKDRKATL 955
Qy 988 NRLFLPKTEFELEHRRFGNTSALDREFFYGLVGRETLIRL-PDVRTPLLVRLDAIS 1046
Db 956 NRLFLPGAREFFPHRASVYGDTSILDSKDFYGLRPGKEYTVLDLPGVR--LLIEQLQAVG 1013
Qy 1047 EPDDKGRNVNVANVGQIRPMVRDRSVESVTATAEKADSSNKGHVAAFPAGVYTVTVAAE 1106
Db 1014 DADSEGRMRTVSSUNGQLRPIQVRDRAATDVPVTEKADRANPCHVAAFPAGVYTVLVAE 1073
Qy 1107 GDEVKAGDAVAITEAMKMEATITASVDGKIERYVVPVPAATKVEGGDLIV 1154
Db 1074 GDEVKAGDAVATITEAMKMEASITAPKSGTVTRLAINRIOQVEGGDLIV 1121

RESULT 4
Q9F843 PRELIMINARY; PRT; 1127 AA.
AC Q9F843;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Pyruvate carboxylase (EC 6.4.1.1).
GN PYC.
OS Mycobacterium smegmatis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_FaxID=1772;
RN [1].
RP SEQUENCE FROM N.A.
RX MEDLINE=20374587; PubMed=10913817;
RA Mukhopadhyay B., Purwantinil E.;
RT "Pyruvate carboxylase from Mycobacterium smegmatis: stabilization,
RT rapid purification, molecular and biochemical characterization and
RT regulation of the cellular level.";
RL Biochim. Biophys. Acta 1475:191-206(2000).
DR EMBL; AF262949; AAC30411.1; -.
DR HSSP; P24182; 1BNC.
DR InterPro; IPR001882; Biotin_attach.
DR InterPro; IPR000089; Biotin_lipoyl.
DR InterPro; IPR000901; CPsase.
DR InterPro; IPR000891; HMGL-like.
DR InterPro; IPR003379; PYC_OADA.
```

```
DR InterPro; IPR000634; S/T_dehydratse.
DR Pfam; PF02785; Biotin_carb_C; 1.
DR Pfam; PF00364; biotin_lipoyl; 1.
DR Pfam; PF00289; CPsase_L_chain; 1.
DR Pfam; PF02786; CPsase_L_D2; 1.
DR Pfam; PF00682; HMGL-like; 1.
DR Pfam; PF02436; PYC_OADA; 1.
DR TIGRFAMS; TIGR01235; pyruv_carbox; 1.
DR PROSITE; PS00188; BIOTIN; UNKNOWN.1.
DR PROSITE; PS00867; CPASE_2; UNKNOWN.1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; UNKNOWN.1.
KW Ligase.
SQ SEQUENCE 1127 AA; 120765 MM; 18C132C48425C67B CRC64;

Query Match 64.6%; Score 3786.5; DB 2; Length 1127;
Best Local Similarity 66.2%; Pred. No. 2.2e-194;
Matches 747; Conservative 142; Mismatches 233; Indels 7; Gaps 6;

Qy 31 KILVANRGIEIVRAPEALETGANTVAIYPREDGSFHRSPFASAVRIGCTESPVKAYLD 90
Db 4 KVLVANRGIEIARFAAYEMGIATVAVIYEDRNLHRLKADESTQIGEGVHPVAYLUS 63
Qy 91 IDEITGAKKVADAIYPOYGFSENAQLARECAENGTIFIGTPEVLDLTGDKSRAYTA 150
Db 64 VDEILRVAKHGADAVIYPOYGFSENPDLAKCAEAGITFFVGPSEVLQLTGNKARATAA 123
Qy 151 AKKAGLPVLAESTPSKNIDDIKVSAGQYPIPVKAVAGGGGRGMRFFVSSPDELRLKATE 210
Db 124 ARAAGLPVLSSSEPSVDELMAAADMEEPLFVKAVSGGGGRGMRVYTDRESLAEALIA 183
Qy 211 ASREAAAFGDSVYVERAVINPOHIEVOILGDRTGVEVHLYERDCSLORRHOKVVEIAP 270
Db 184 ASREAAEFAGDSVYLEQAVLNPHIEVOLADGAGNVNHLFERDCSVORRHOKVVEIAP 243
Qy 271 AGHLOPELDRDICADAVKFCRSIGYQAGTVFELVDEKGNHVFIEIMNPRIOVEHTVTEEV 330
Db 244 APNLSDELQQICADAVAFARQIGYSCAGTVFELDERGHHVFIECNPRIOVEHTVTEEI 303
Qy 331 TEVDVLKAMRLAAGATLKELGLTQDKIKHGAALQCRTTDPNNGRFPDGTGTAYRS 390
Db 304 TDVDLVSSOLRIAAGETLADLGSQDLRVVGAAMQCRTTTEVPANGFRPDGTGRITAYRS 363
Qy 391 PGAGVRLDGAALGGEITAHFDSMLVKMTCRGSDFAVARAQRALAEFTVSGVATNIG 450
Db 364 PGAGVRLDGGTNLGAEISAHFDSMLVKLTCRGDRFSAASARRALAEFRIGVSTNIP 423
Qy 451 FLRALLREEDFTSKRIATGFTGDHPHLLQAPPADDEOGRILDYLDVTVNKPVGPRKDV 510
Db 424 FLQAVIDDPDFRAGRVTTSFIDDRPHLLTSRSPADRGTRILNVLADITVKNKPHGERPSTV 483
Qy 511 AAPIDKLPNIKDL--PLPGSRDRLLKQLGPAFAFARDLREDDALAVTDTTFRDAHQSLIAT 568
Db 484 -YQDKRLPPL-DLQAPPAGSKRLVELGEGFAGWLRESKAVGVTDTTFRDAHQSLIAT 541
Qy 569 RYRSFALKPAAEAVAKLTPELLSVEAWGATYDVAMRFLFEDPDWDLDELREAMPNVIQ 628
Db 542 RYRTTGLMVAIPYVARSMFOLLISICWGATYDVALLFLKEDPWERLALARESVFNICLQ 601
Qy 629 MLLRGRNTVGYTPYDPSVCRFAFYKEAASSGVDFIRIFDALNDVDSQMRPAIDAVLETNTAV 688
Db 602 MLLRGRNTVGYTPYPELVTSFAFVEEAATGIDIFRIFDALNNVESWRPAIDAVRETGSTI 661
Qy 689 AEVAMAYSGDLSDPNEKLYTLDYLLKMAEIVKSGAHILAIKDMAGLLRPAAVTKLVTA 748
Db 662 AEVAMCYTGDLSDPAENLYTLDYLLKLAEQIVEAGAHVLAIKDMAGLLRPAAPATLVA 721
Qy 749 RREFDLPVHVHTDTHAGGOLATYFAAAQAGADAVDCASAPLSGTTSSQPSLSAIVAAFAHT 808
Db 722 RSRFDLPVHVHTDTHPGGOLATYLAAWSAGADAVDCASAPMAGTTSSQPSLSAIVAAFAHT 781
Qy 809 RRDGTGLSLEAVSDLEPYWEAVRGLYLPFESGTPGPRGVYRHEIPCGOLSNLRACATL 868
Db 782 QYDTCGLDLRAVCDLEPYWEAVRKYAPFESGLPGTGRVYTHEIFCGQLSNLRQAALG 841
```

QY 869 LADRELEIDNYAANEMGLRPTKVTPTSSKVVGDIALHLVAGVDPADFAADPQKYDIPD 928  
 DB 842 LGRDEEIEIANYAADRVLGRVKTPTSSKVVGDIALALVAGITAEFAEDPAKYDIPD 901  
 QY 929 SVIALRGLGNPPGGWPEPLRTRALEGRSEKAPLTPVEPEQAHLDADDSKERNLSN 988  
 DB 902 SVIGLRLGELGPPGGWPEPLRTKALQGRGPAP-PVEKUTADDEALL-AQPGPKRQAALN 959  
 QY 989 RLIFPKPTPEEFLEHRRFCNTSALDDREFYGLVCEGRETLIRLPDVRTPLVRLDAISEP 1048  
 DB 960 RLIFPGPTAEFAHRETYGDTSSANOFFYGLRYGEEHRVQL-ERGVELLIGLEAISEA 1018  
 QY 1049 DDKGRNVVANYNGQIRMRVDRDSVESVTATAEKADSNKGHVAPFAGVTVTVTAEGD 1108  
 DB 1019 DERGMRTVMCIINGQLRPVLVRDRDSIASEVPAEAKADRNADHIAAPFAGVTVTVTAEGD 1078  
 QY 1109 EVKAGDAVAILIEMKMEATITASVCGKIERVVVPAATKVEGGDLVVVS 1157  
 DB 1079 SVDAGOTIATIEAMKMEAITAPKAGTVARVAVATAQVEGGDLVVVS 1127

RESULT 5  
 P95127  
 ID AC P95127 PRELIMINARY; PRT: 1127 AA.  
 DT 01-MAY-1997 (TReMBLrel. 03, Created)  
 DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Hypothetical 120.4 kDa protein (Pyruvate carboxylase).  
 OS PCA OR RV2967C OR MTCY349.20 OR MT3045.  
 GN Mycobacterium tuberculosis.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_taxid=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekai F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.:  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / OSHKOSH;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; 283018; CAB05410.1; -.  
 DR EMBL; AE007125; AAK47371.1; -.  
 DR HSP; P24182; 1BNC.  
 DR TIGR; MT3045; -.  
 DR TubercuList; RV2967c; -.  
 DR InterPro; IPR001882; Biotin\_attapch.  
 DR InterPro; IPR000089; Biotin\_lipoyl.  
 DR InterPro; IPR000901; CPSase.  
 DR InterPro; IPR000891; HMG-Like.  
 DR InterPro; IPR003379; PYC\_OADA.  
 DR InterPro; IPR000634; S/T\_dehydrtse.  
 DR pfam; PF02785; Biotin\_carb\_C; 1.

DR pfam; PF00364; biotin\_lipoyl; 1.  
 DR pfam; PF00289; CPSase\_L.Chain; 1.  
 DR pfam; PF02786; CPSase\_L\_D2; 1.  
 DR pfam; PF00682; HMG-Like; 1.  
 DR pfam; PF02436; PYC\_OADA; 1.  
 DR TIGRFAMS; TIGR01235; pyruv\_carbox; 1.  
 DR PROSITE; PS00188; BIOTIN; UNKNOWN.1.  
 DR PROSITE; PS00867; CPSASE.2; UNKNOWN.1.  
 DR PROSITE; PS00165; DEHYDRATASE\_SER\_THR; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 1127 AA; 120422 MW; 84B0A4CC1A23CD90 CRC64;

Query Match 63.3%; Score 3713.5; DB 16; Length 1127;  
 Best Local Similarity 64.4%; Pred. No. 1.8e-190;  
 Matches 730; Conservative 153; Mismatches 239; Indels 11; Gaps 6;

QY 29 FKKILVANGETAVRAFAALETGAATVAIYPREDRGSHRSFASAEVRIGTGEGSPVKAY 88  
 DB 2 FSKVLVANGETAVRAFAAYELGVGTVAIYPVEDRNSQHRLKADESQYIGDIGHPVAY 61  
 QY 89 LDIDEILGAAGKVKADAIYPGYGFLSENALARECAENGITFETGPTPEVLDTGTGDSRAV 148  
 DB 62 LSVDEIVATARRAGADAIYPGYGFLSENPDLAACAAAGISFVGPSEAEVLELAGNSRAI 121  
 QY 149 TAAKAGLPVLAESTPSKNIDDIIVKSAEGQTYPIFYKAVAGGGGRMRFVSSPDRLKLA 208  
 DB 122 AAAREAGLPVLMSSAPSASVDELLSVAAGMPPLFLFKAVAGGGGRMRRVQDIATLPEAI 181  
 QY 209 TEASRAEAAFGSGSVYVERAVINPOHIEVQILGDRTEGVVHLYERDCSLQRHRQKWEI 268  
 DB 182 EASREAESAFGDPVTYLEQAVINPRHIEVQILADNLGVDVHLYERDCSVQRRHRQKVI 241  
 QY 269 APAQHLDPELRICADAVKFCRSIGYCGAGTVEFLVDEKGNHVFTEMNPRIOVEHTVTE 328  
 DB 242 APAPHLDIAELRYKMCVDVAFARHIGYSCAGTVEFLVDEKGEYVFTENPRVQVEHTVTE 301  
 QY 329 EVTEVDLVKAQMLAAGATLKEGLTQDKIKTHGAALQCRITTFEDPNNGFRPDGTITAY 388  
 DB 302 EITDVDLVASQRIAGETLEQLGRQEDIAHPGAALQCRITTEDPANGFRPDGTGRISAL 361  
 QY 389 RFGGAGVRLDGAALGGEITAHFDSMLVMKTCRGSDFTAVARAQALAEFTVSGVATN 448  
 DB 362 RTAGGAGVRLDGSNLGAEISPYFDSMLVKLTCRGDLPTAVSARRAIAEFTIRGVSTN 421  
 QY 449 IGFLRALLREEDTSKRIATGFTGDHPHLLQAPPADDEQGRILDLADVTNPKPHGVPRK 508  
 DB 422 IFLQAVLDPPDFRAGRVTTSIFDERPQLLTARASADRGTKILNLFADVTNNPVSRRPS 481  
 QY 509 DVAAPIDKLPNIKDEPL----PRGSRDLKQLGPAAFAFARDLREQDALAVTDTTFRDAHQ 564  
 DB 482 TI-YPDCKLP---DLDLRAAPPAGSKQRLVKLGPGEFARWLRESAAVGVDTTFRDAHQ 537  
 QY 565 LIATRVRSFALKPAAEAVAKLTPELLSVWAGGATYDVAMRFLFDPDPWDLDELREAMPN 624  
 DB 538 LIATRVTSGLSRVAPYLAARTPQLLSVWCGGATYDVVALRFELKEDPWERLATLRAAMPN 597  
 QY 625 VNTQMLLRGNTVGYTPYPSVCRAFEVKBAASSGVDFIERIFDALNDVDSOMRPAIDAVLET 684  
 DB 598 ICLQMLLRGNTVGYTPYPIEYTSFVQATATGIDGIDIFRIDALNNIESMRPAIDAVRET 657  
 QY 685 NTAAVEAVMAYSGDLPNEKLYTLDYLYLKMAEIEVKSGAHILAIKDMAGLLRPAAVTKL 744  
 DB 658 GSAIAEAVMAYTCDLTDPCGELYTLDYLYLKLAEQIVDAGAHVLAIKDMAGLLRPAARQL 717  
 QY 745 VTALRREFDLPVHVVHTDTAGGOLATYFAAQAGADAVGASAPLSGTTSQPSLSAIVAA 804  
 DB 718 VSALRSRFDLPVHLHTDTPGGOLASYAAWHAGADAVDGAAPLACTTSQPSLSIVAA 777  
 QY 805 FAHTRDTGLSLAEAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGQSLSNRAQA 864  
 DB 778 AAHTEYDTGLSLSAVCALEPYWEALRKVYAPFESSLPGPTGRVYRHEIPGQSLSNRQQA 837  
 QY 865 TALGLADREFELIDNVAANEMGLRPTKVTPTSSKVVGDIALHLVAGVDPADFAADPQKY 924

```
Db 838 IALGLDRFEDIEEAYAGADKVLGRUVKVTPTSKVVGDLALALVAGVSADEFASDPARF 897
Qy 925 DIPDSVIAFLRGLNGELNPPGWPPELRLTRALEGRSEKAPLTPVEPEEQAHLDDADSKERR 984
Db 898 GIPESVLGFLRGLGDPGGWPELRLTRALAGRAAR-PTAQAADDEIALSSVGAK-RQ 955
Qy 985 NSLNRLLFPKPTPEFLHRRFRFGNTSALDDREFFYGLVGEHRETLIRLPDVRTPLLVRLDA 1044
Db 956 ATLNRLLFPSPKTEFNEHREAYAGDTSOLSANQFFYGLRQGEHREVKL-ERGVELLGLEA 1014
Qy 1045 ISEPDGMRNVANVNGQLRPMRVRDRSVESVTATAEKADSNKNGHVAAPFAGVVTVT 1104
Db 1015 ISEPDGMRVTMVCILNQLRPLVLRDRSISASVPAEAKADGNPGHIAAPFAGVVTGV 1074
Qy 1105 AEGDEVKAGDAVATIEAMKWEATITASVCGKIERVVVPAATKVEGDLIVVVS 1157
Db 1075 CVGERVAGGTIATIEAMKWEAPITAPVAGTVVERVAVSDTAQVEGDLIVVVS 1127

RESULT 6
ID Q50450 PRELIMINARY; PRT; 1124 AA.
AC Q50450;
DF 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
PYC.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OX Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Robison K.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
EMBL; U00024; AAA50948.1; -.
DR HSP; P24182; IBNC.
DR InterPro; IPR001892; Biotin_attach.
DR InterPro; IPR000089; Biotin_lipoyl.
DR InterPro; IPR000901; CPsase.
DR InterPro; IPR000891; HMGL-like.
DR InterPro; IPR003379; PYC-OADA.
DR InterPro; IPR000634; S/T_dehydratse.
DR Pfam; PF02785; Biotin_carb_C; 1.
DR Pfam; PF00364; biotin_lipoyl; 1.
DR Pfam; PF00289; CPsase_L; 1.
DR Pfam; PF02786; CPsase_L-D2; 1.
DR Pfam; PF00682; HMGL-like; 1.
DR Pfam; PF02436; PYC-OADA; 1.
DR TIGRFAMS; TIGR01235; pyruv_carbox; 1.
DR PROSITE; PS00188; BIOTIN; UNKNOWN.1.
DR PROSITE; PS00867; CPsase_2; UNKNOWN.1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
SQ SEQUENCE 1124 AA; 120672 MW; 07E1692E12203491 CRC64;

Query Match 59.4%; Score 3484; DB 2; Length 1124;
Best Local Similarity 62.2%; Pred. No. 3 5e-178;
Matches 698; Conservative 151; Mismatches 251; Indels 22; Gaps 9;

Qy 29 FKKTLVANRGEIARAFRALETCATVATYPRDRGSPHRSFASAVRIGTEGSPVKAY 88
Db 2 FSKVLVANRGEIARAFRALETCATVATYPRDRGSPHRSFASAVRIGTEGSPVKAY 61

Qy 89 LDTDEITGAKKKADAIYCYGFLSENAQALARECAENGTITFGPTPEVLDTGDKSRAY 148
Db 62 LSVDEIVATARRAGADAIYCYGFLSENPDLAAACAAAGISFVGPSEAULELAGNSKRAI 121

Qy 149 TAAKAGLPVLAESTPSKNIDIVKSAEGQTYPIFKAVAGGGGGRMRFVSSPDELKLA 208
```

```
Db 122 AAAREAGLPVLMSSAPSASVDELLSVAAGMPPELFVKAVAGGGGMRVRDIAALPEAI 181
Qy 209 TEASREAAAFGDSGVYVERAVINPOHIEVQILGORTGEVHLYERDCSLQRRHKQVVEI 268
Db 182 EAASREASAFGDPVTYLEQAVINPRHIEVQILADNLGDIHLYERDCSVQRHQKVIETL 241
Qy 269 APAQHLDPRLRORICADAVKFCRSIGYOCAGVFEVLDEKGNHVIEMNPRTOVHTYTE 328
Db 242 APAPHDLAELRYKCMKDAVAFARHIGYSCAGTVEFLDERGEYVEIEMNPRVOVHTYTE 301
Qy 329 EYTEVDLVKAOMRLAAGATKELGLTODKIKTHGAALOCRITTEDPNNGFRDGTGTAY 388
Db 302 EITVDLVASQIRIAGETLEQGLRQEDIAFHGAALOCRITTEDPANGFRP-TRAGSAR 360
Qy 389 RSPGGAGVRLDCAAOGLGEITAHFOSMLVKM-----TCRGSOFETAVARAQALAEFTVS 443
Db 361 COPPAVPVSAMTAAP-----TWRRNOPVLRHAGQADLSGRDLPYAVSHARRAIEFRIR 415
Qy 444 GVATNIGFLRALLREEDFTSKRIATGFIGDHPHLIQAAPPADDEOGRILDYLAADVTVNKP 503
Db 416 GVSTNIPFLQAVLDPPDFRAGRVTTSFIDERPOLLTARASADRGTKILNFLADVTVNNPY 475
Qy 504 GVRPKDVAAPIDKLNKIDPL-----PRGSRDLKQLGPAFAFARDLREQDALADLTDTTFR 559
Db 476 GSRPSTI-YPPDKLP---DLDLRAAPPAGSKQRLVKLGPEGFARWLRESAAVGVDTTFR 531
Qy 560 DAHQSLLATRVRSFALKPAAEAVALTPELISVEAWGATYDVAMRFLFEDPDWRLDELRL 619
Db 532 DAHQSLLATRVRTSGLSRVAPYLARTMPOLLSVECWGGATYDVALRFLKEDPWERLATLR 591
Qy 620 EAMPNVNIQMLRGRNTVGYTPYPSVCRAFVKEAASGVDFIRFDALNDVVSOMRPAID 679
Db 592 AAMPNICLOMLLRGRNTVGYTPYPEIVTSAFVQEATATGIDIFRFDALNINIESRPAID 651
Qy 680 AVLENTVAEAVAMAYSGDSDPNKLYTDLYLKMAEIVKSGAHILAIDKAGLLRPA 739
Db 652 AVRETSALAEVAMCYTGDLIDPGELTYLDLYLKAGQIVDGAHVLAIDKAGLLRPP 711
Qy 740 AVTKLVTLALREFDLPVHVHTHDTAGGOLATYFAAAGADAGADVAGASAPLSTTSQPSLS 799
Db 712 AAQRLVSALRSRFDLPVHLHDTGPGQLASYAAAHAGADAVDGAAPLACTTSQPALIS 771
Qy 800 AIVAAFAHTRDTGLSLEAVSDLEPYWEAVRGLYLPFSGTPGTPGRVYRHEIPGGQLSN 859
Db 772 SIVAAAHTXYDTGLSLSAVCALEPYWEALRKVYAPFESGLPGTPGRVYHHEIPGGQLSN 831
Qy 860 LRAQATAGLADRFELEDNYAAVNEMLGRPTKVTSSKVVGDLALHLVAGVDPADEFAA 919
Db 832 LRQQAIALGLGRFEEIEEAYAGADRVLRGVKVTPTSKVVGDLALALVAGVSADEFAS 891
Qy 920 DPQKYDIPDSVIAFLRGLNGELNPPGWPPELRLTRALEGRSEKAPLTPVEPEEQAHLDD 979
Db 892 DPARTGIPESVLGFLRGLGDPGGWPELRLTRALAGRAAR-PTAQAADDEIALSSVG 950
Qy 980 SKERNLSNRLLPKPTPEFLHRRFRFGNTSALDDREFFYGLVGEHRETLIRLPDVRTPL 1039
Db 951 AK-RQATLNRLLFPSPKTEFNEHREAYAGDTSOLSANQFFYGLRQGEHREVKL-ERGV 1008
Qy 1040 VRLDAISPEDDKMRNVANVNGQLRPMRVRDRSVESVTATAEKADSNKNGHVAAPFAGV 1099
Db 1009 IGLEAISPEDEKMRVTMVCILNQLRPLVLRDRSISASVPAEAKADGNPGHIAAPFAGV 1068
Qy 1100 VTVTVAEGDEVKAGDAVATIEAMKWEATITASVCGKIERVVV 1141
Db 1069 VTVGVGVGERVAGGTIATIEAMKWEAPITAPVAGTVVERVAV 1110

RESULT 7
Qy 9K9M0 PRELIMINARY; PRT; 1150 AA.
AC Q9K9M0;
DT 01-OCT-2000 (Tremblrel. 15, Created)
```

DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)	Db	484	GYPGLE-KTKKPVFDKPPVKLKLSEIPDGTQKILDOHGPEGLAKWVKQKQVLLTDTT	542
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)	Qy	558	FRDHOSLLATRVRSFALKPAAEAVALKTPELLSSVEAWGATYDVANRFLFEDPWRDLE	617
DE	Pyruvate carboxylase.	Db	543	FRDHOSLLATRVKTHDLKQIAEPTARLLPNLFMSMEMMGATFDVAMRFLHEDPWERLLI	602
GN	PYCA OR BH2625.	Qy	618	LREAMPNVNIQMLLRGRNTYPTYPDSVCRAFFVKEAASGVDFRIFDLDNDVDSMRPA	677
OS	Bacillus halodurans.	Db	603	LRKKAPNVLFOMLLRASNAVGYKNYPDLNREFVYDKSANAGIDVFRIFDLSLNVGEMKLA	662
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;	Qy	678	IDAVLENTAVAEVAMAYSGDLSDPNEKLYTLDYLLKMAEEIVKSGAHILAIDMAGLLR	737
OC	Bacillaceae; Bacillus.	Db	663	IEAVGEAN-KIAEATICYTGDIILSSRPKYLAYKKLAKELAAGAHILGIDMAGLLK	721
OX	NCBI_TaxID=86665;	Qy	738	PAAVTKLVTLRRFEDLPVHVHTDAGGOLATYFAAAQAGADAVDCASAPLSGTTTQPS	797
RN	SEQUENCE FROM N.A.	Db	722	PEAYQLVLAELKDTVTIPVHLHTDTHSGNGIFTIYARIEAGVDIVDVAVSSMAGLTQPS	781
RC	STRAIN=C-125 / JCM 9153;	Qy	798	LSAIVAAFAHTRRDGTGLSLEAVSDLEPYWEAVRGLYLPFESGTPGPTGRVYRHEIPGQL	857
RX	MEDLINE=20512592; PubMed=11058132;	Db	782	ANSLYVALADSERQPNVNIITALEQLAEFWEETRKFYAGFESGMNAPHTVEYHEMPGQY	841
RA	Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,	Qy	858	SNLRQAATGLADRFELIEDNYAANVEMLGRPTKVTTPSSKVVGDLALHLVGACVDPADF	917
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,	Db	842	SNLQQAQAKAVGLGRWNEVKMYRTVDMFGDVVKVTPSSKVVGDMALYVQNDLTBEV	901
RA	Horikoshi K.;	Qy	918	AADPOKYDIPDSVIAFLRGLGNPPGCGPEPLRTRALEGRSEKAPLTVPEEEOAHLDA	977
RT	"Complete genome sequence of the alkaliphilic bacterium Bacillus	Db	902	YENGHLKDFDPSVVEFEGQLGYPQGFKKLOEIIILKGRK---PITNRPGENNEPIQF	957
RT	halodurans and genomic sequence comparison with Bacillus subtilis.";	Qy	978	DDSKER-RNSLNR-----LLFPKPTTEEFLEHRRFRGNTSALDRDFEYFGLVGE	1026
RL	Nucleic Acids Res. 28:4317-4331(2000).	Db	958	EAIKEELYNKLDQVTSHTDILSYALPKVMEFERPQTGDSVSLDTPTFYGLRPGEE	1017
DR	HMBL; AP001516; BAB06344.1; -.	Qy	1027	TLRLDPVRTPLVRLDAISEPDDKGMNVNVANVNGOIRPMRVDRDSVESVTAEKADS	1086
DR	InterPro: IPR000089; Biotin_lipoyl.	Db	1018	IEVEIEOGKT-LIVKFISLSKPDQDGNRIYFELNGOPREVLINDQSVKTSIISRPKADK	1076
DR	InterPro: IPR000901; CPase.	Qy	1087	SNKGHVAAPFAG-VVTVTVAEGDEVKAGDAVAIIEMKMEATITASVDGIKRVVVPAA	1145
DR	InterPro: IPR000891; HMGL-like.	Db	1077	SNPHNGASMPGIVVVKALVEKDKVKQGDHLMTEAMKMETTVQAPDGEVVALHVKGDG	1136
DR	InterPro: IPR003379; PYC_OADA.	Qy	1146	KVEGGDLIVVV 1156	
DR	Pfam: PF02785; Biotin_carb_C; 1.	Db	1137	AIQTGDLLEIV 1147	
DR	Pfam: PF00364; biotin_lipoyl; 1.				
DR	Pfam: PF00289; CPase_L_chain; 1.				
DR	Pfam: PF02786; CPase_L_D2; 1.				
DR	Pfam: PF00682; HMGL-like; 1.				
DR	Pfam: PF02432; PYC_OADA; 1.				
DR	TIGRFAMS; TIGR01235; pyruv_carbox; 1.				
DR	PROSITE; PS00867; CPASE_2; UNKNOWN_1.				
KW	Complete proteome.				
SW	SEQUENCE 1150 AA; 128858 MW; ED6788BB8A9F3BD4 CRC64;				
Query Match 43.2%; Score 2535.5; DB 16; Length 1150;					
Best Local Similarity 46.9%; Pred. No. 2.4e-127;					
Matches 540; Conservative 185; Mismatches 399; Indels 27; Gaps 13;					
Qy	26 LPAFKKILVANRGEIAVRAFALEATCAATVAIYPREDGSRFSRASEAVRIGTEGSPV	85			
Db	4 LKNIKVLVANRGEIAIRIFRATCELHIRTVAIYSKEDTGAYHRYADEAYLVGEGKKPI	63			
Qy	86 KAYLDEIETGAIAKKVADAIYPGYFLSENAQLARECAENGITFGPTPEVLDTGDKS	145			
Db	64 EAYLDIEGIIETIAKRGVDAIHGPGYFLSENFIAKRCHEEIGFIQPELHLVMEGDKV	123			
Qy	146 RAVTAKKAGLPVLAEST-PSKNIDDIVKSAEQCTPIFVKAVAGGGGGRMRVSSPEL	204			
Db	124 QAREQAIKANLPVIGSDGVSLEDDYKAFADKHGYPFIKAALGGGGGRMIVRSENV	183			
Qy	205 RKLATEASREAAAGDGVYVVRVAVINPOHIEVQILGDRTEGVVHLRYDRDCSLORRHQK	264			
Db	184 QESYERAKSEAKAFNGDENVEYVEKFTENPKHIEVQILADKHGNTLHLHYERDCSVQRHQK	243			
Qy	265 VVEIAPAHLDELDRICADAVKFCRSIGYOGAGTVEFLVDEKGNHVFTEMNPRIQVEH	324			
Db	244 VVEVAPSVLSDEVRIQCQAQVLAENVYVNAVGTVEFLVDREGNFYFIEVNPRIQVEH	303			
Qy	325 TVTEEVTEVDLVKAQMRLAAGATL--KELGL-TQDKIKTHGAALQCRITTEDPNNGRFPD	381			
Db	304 TITETWTGIDIVQSOLFADGEHLHGRDLPKQEEIVCHGYAIQSRVTTEDPSNGFLPD	363			
Qy	382 TGTITAYRSPGGAGVRLD-GAAQLGGEITAHFDSMLVKMTCRGSDPETAVARAQALAEF	440			
Db	364 TGRINAYRSGGGFVRDLADAGNGFQGAIVTIPYDLSLVKYSTWALTTEGAARKMLNRLEF	423			
Qy	441 TVSGVATNTGFRALLREEDFTSKRTATGFIGDHPHLLQAPADDEQGRILDYLDVTVN	500			
Db	424 RIRGIKNTIAFLNVVQHQRFLSGEYNTSFIDQTPELFVPPKRDCTKMLSFIGETIVN	483			
Qy	501 KPHGVPRKDVAAPIDK--LPNIK-DLPLPRGSRDLKQLGPAFAFARDLREQDALAVTDTT	557			









```
DR InterPro: IPR001882; Biotin_attach.  
DR InterPro: IPR000089; Biotin_lipoyl.  
DR InterPro: IPR000901; CPSase.  
DR InterPro: IPR000891; HMGL-like.  
DR InterPro: IPR003379; PYC_OADA.  
DR Pfam: PF02785; Biotin_carb_C; 1.  
DR Pfam: PF00364; biotin_lipoyl; 1.  
DR Pfam: PF00289; CPSase_L_chain; 1.  
DR Pfam: PF02786; CPSase_L_D2; 1.  
DR Pfam: PF00682; HMGL-like; 1.  
DR Pfam: PF02436; PYC_OADA; 1.  
DR TIGRFAMs: TIGR01235; pyruv_carbox; 1.  
DR PROSITE: PS00188; Biotin; FALSE_NEG.  
DR PROSITE: PS00866; CPSASE_1; 1.  
DR PROSITE: PS00867; CPSASE_2; UNKNOWN_1.  
DR Ligase; Multifunctional enzyme; Biotin; Gluconeogenesis; ATP-binding.  
DR FT DOMAIN 1 520 BIOTIN CARBOXYLASE (BY SIMILARITY).  
DR FT DOMAIN 521 968 CARBOXYLTRANSFERASE (BY SIMILARITY).  
DR FT DOMAIN 1064 1147 BIOTIN CARBOXYL CARRIER PROTEIN  
      (BY SIMILARITY).  
DR NP_BIND 167 172 ATP (BY SIMILARITY).  
DR ACT_SITE 296 296 BY SIMILARITY.  
DR FT BINDING 1112 1112 BIOTIN (BY SIMILARITY).  
DR SEQUENCE 1147 AA; 128336 MW; DFD2F003F838F591 CRC64;  
Query Match 42.8%; Score 2509; DB 2; Length 1147;  
Best Local Similarity 46.8%; Pred. No. 6.4e-126;  
Matches 539; Conservative 181; Mismatches 397; Indels 34; Gaps 15;  
QY 30 KKILVANGETAVAFRAALETGAATVAIYPRDRGCSFHRSFASEAVRIGTEGSPKAYL 89  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
7 RKVLVANGELAIKRVFRACTELGIHRTVAIYSKEDVGHRYKADAEVLYGEGKKPIEAYL 66  
QY 90 DIDEIIGAANKVKADAIYGYGFLSENAQLARECAENGITGFTPTPEVLDTGDKSRVAT 149  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
67 DIEGIEIAKADVADAIHPGQFSENIQAKKREGEIIFPGPNENHLMDFGDKVKARH 126  
QY 150 AAKKAGLPVLAEST-PSKNIDDIIVKSAEQTYPIFVKAVAGGGRGMRVFSVSPDELRKLA 208  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
127 AAVNAGIPVPGSDGVGDVLEVVAFAEAHGYPIIKAAALGGGGGMRGFRVRSKSEVKEAF 186  
QY 209 TEASREAAAFGSDGVSVERAVINPOHIEVOILGDRDTGEVYVHLHYERDCSLQRHQKWEI 268  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
187 ERAKSEAAAFGSDGVYVYKLIENPKHIEVQILGDYEGNTVHLHYERDCSVQRHQKWEV 246  
QY 269 APAQHLDELDRICADAVKFRCSIGYOGAGTEFFLV--DEKGNHVFIEKNPRIQVHEVT 326  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
247 APSVSLDELQRICEAAVOLMRSGVYVNACTVEFLVSGDE--FVPIENVPRIQVHEITI 303  
QY 327 TEEVTEVDLVKAQMLAAGATL--KELGL-TQDKIKTHGAALQCRITTEDPNNGFRPDTG 383  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
304 TEMITGIDIVQSQRLIADGCSLHSEVGIKQEDIRINGVYAIQSVRTTEDPLNFMPTDG 363  
QY 384 TITAYRPGGAGVRD--GAAQLGGEITAHFDSMLVKMTCRGSDFETAVARAQALAEFTV 442  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
364 KIMAYRSGGGFGVRLDAGNFGOGAVITPYVDSLVLKSLTWTALTFEQAARKMLNLRTERI 423  
QY 443 SGVATNIGFRLALLREEDFTSKRIATGFIGDHPHLLQAPPADDEQGRILDYLDADVTNKP 502  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
424 RVFKTNIPLENVQHPKFLSGEYDTSFIDTTELVFPFRKDKGKMLTYIGTVVNGF 483  
QY 503 HGVRPKDVAAPI---DKLPNKKDL-PLPRGSRDLKLGFAAROLDREODALAVTDTF 558  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
484 PGICKK--KKPEVDKPLRPLKSEAPIGACTKQILDKHGEGVLVRIQEPVLLIDTTF 541  
QY 559 RDAHOSLLATVRFSALKPAEAVAKITPELLSVKAGGATYDVANRFLFEDPDWDELDEL 618  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
542 RDAHOSLLATVRIVDLVRAEFSARLLPNLFLEMMGGGATFVAFVFLKEDPDWDLRLKL 601  
QY 619 REAMPNWNIOMLLRNTVGYTTPYSDSVCFRAVKEAASSGVDFRIFDALNDVDSQMRPAI 678  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
602 RDAFPNVLFOMLLRKSANAVGKKNPDNVITREFEKSAAHAGIHVFRIFDLSLNWVKGMTVAI 661
```

```
QY 679 DAVLENTAFAEVAMAYSGLDSPNEKLYTLDYLLKMAEEIYKSGAHLIAIKDMAGLLRP 738  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
662 DAVROSG-KIAEAAICYGTGDIIDPSRSKYNDLYYKALAKELEQAGAHILAIKDMAGLLKP 720  
QY 739 AAVTKLVLTALRREFDLPHVHTHTAGGLATYFAAAQAGADAVDGASAPLSGTTSQPSL 798  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
721 QAAHVLSLGLKETVDIPIYLHTHTSGNGIYTYAKAIEAGVDIVDVAISSMAGLTSQPSA 780  
QY 799 SAIVAAFAHTRRDTGLSLEAVSDLEPPYWEAVRGLYLPFESGTPGPTGVRVYRHEIPGGOLS 858  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
781 NTLVYALECTERAPEVDIYGLEQLARYWEDVRKFQEFESGNMHPTEVYMHMPGGQYS 840  
QY 859 NLRAQATALGLADRPFLIEDNVAANVMGLRPTKVTYPSKVVGDALHLVAGVADPADFA 918  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
841 NLQQAQAKVGLGDRWDEKEMRYRNDLFGDIVKVTYPSKVVGDALYMYVQNNLTQDDIF 900  
QY 919 ADPOKYDIPDSVITAFRLGELGNPPGCGPEPLTRALEGSECKAPLTVPEPEQAHLAD 978  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
901 ERGETLNFDSVVVELPEGYLGPHGFPKELQRIILKGRE----PITVPPGELLEPPYDE 956  
QY 979 D-SKERNLSNR-----LLFPKPTTEFLHRRRGNTSALDDREFFVGLVEGRET 1027  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
957 QMKKELYDKLGREVTFDFAIAIYALPKVFEYAEVKEVYGVSVLDTPFTFLGMLGEBI 1016  
QY 1028 LIRLPDVRTPLVRLDAISEPDDKGRNVVANNVQIRPMRVDRSVSVTATAEKADSS 1087  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
1017 EWEIERGKT-LIVKLVSIGQPADGTRVYFELNGQPREVIRDESIKTAVVVEHTRADRT 1075  
QY 1088 NKGHVAAFPAG--VVTVYVAEGDEVKAGDAVAILIEMKMEATITASVDGKIERYVVPAAATK 1146  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
1076 NPNHIAATMPGTVKVLVEKGEKVDKGDHLMVTEAMKMETTVQAPFAGVVKDIYVKSQDA 1135  
QY 1147 VEGGDLIVVVS 1157  
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
1136 IOAGDLLMELS 1146  
RESULT 11  
QYHES8  
ID QYHES8 PRELIMINARY; PRT: 1192 AA.  
AC QYHES8;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Pyruvate carboxylase (EC 6.4.1.1).  
GN PYC.  
OS Aspergillus niger.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
OX NCBI_TaxID=5061;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=N400;  
RA Panneman H., Ruijter G.J.G., Van den Broeck H.C., Visser J.;  
RT "Aspergillus niger pyruvate carboxylase";  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
CC -|- COPACTOR: BIOTIN (BY SIMILARITY).  
DR EMBL: AJ009972; CAC19838.1; --  
DR HSSP: P24182; 1BNC  
DR InterPro: IPR001882; Biotin_attach.  
DR InterPro: IPR000089; Biotin_lipoyl.  
DR InterPro: IPR000901; CPSase.  
DR InterPro: IPR000891; HMGL-like.  
DR InterPro: IPR003379; PYC_OADA.  
DR Pfam: PF02785; Biotin_carb_C; 1.  
DR Pfam: PF00364; biotin_lipoyl; 1.  
DR Pfam: PF00289; CPSase_L_chain; 1.  
DR Pfam: PF02786; CPSase_L_D2; 1.  
DR Pfam: PF00682; HMGL-like; 1.  
DR Pfam: PF02436; PYC_OADA; 1.  
DR PRINTS: PR00098; CPSASE.  
DR TIGRFAMs: TIGR01235; pyruv_carbox; 1.  
DR PROSITE: PS00188; Biotin; 1.
```

Qy	30	KKILVANRGEIAVRAALETGAATVAIYPREDRGSGFHRSFASEAVRIGTEGSPVKAYL 89
		:  :
Db	5	KKVLVANRGEIAIRVMRACTELIKTKTVAIYSQEDTGSHFYKSDEAYLVGAGKKPIDAYL 64

```
QY 90 DIDEIIGAARKVADAIYPGYGFLSENAQLARECAENGITFIGTPEVLDLTGDKSRVAT 149
DB 65 DIENIELAKESGADAIHPGYGFLSENFARFARCEQEGIIIFVGPCKSHLDMFGDKIKAKE 124
QY 150 AAKKAGLPVLAEST-PSKNIDDIIVKSGQOTPIFVAVAGGGRGMRFVSSPDELKLA 208
DB 125 QALLADIPVPGSNGPVAGIKEVEEFGKNGYPLMIKASLGSGGRGMRFVSKHEHVRESF 184
QY 209 TEASREAAAGDCSVYVERAVINPOHTEVOILGDRTEGVVHLVYERDCSLORRHOKVVEI 268
DB 185 ERASSEAKAAGNDEVYVEKCMVMPKHIEVOILGDTGNIVHLFERDCSLORRHOKVVEV 244
QY 269 APAQHLPELDRICADAVPCRSIGYGAGTVEFLVDEKGNHVFIEEMNPRIQVEHTVTE 328
DB 245 APCNAITSLSNRNICDAVKMLKMWYINAGTVEFLV-EGDDFYFIEVNPVQVEHTITE 303
QY 329 EVTEVDLVKAOMLAAGATIKELGLT---QDKIKTHGAALQCRITTEDPNNGFRPDGTGI 385
DB 304 MITGIDIVQSOLFADGYALHQDVAIAPKQEDIHGSAIQSRITTEDPLNFMPTDTRGV 363
QY 386 TAYRSPGGAGVRLD-GAOLGGEITAHFDSMLVKMTCRGSDFETAVARAQALAEFTVSG 444
DB 364 DTYRSTGGVRLDAGNGFQCTVVTPEYDSLLVKLTGWTGTFEQATKMRNLTIEFRIG 423
QY 445 VATNIGFLRALLREDEFTSKRIATGFTGDHPHLLQAPPADDEQGRILDYLDVTVNKPFG 504
DB 424 VKTNIPELLNVVRHDPDFASNGYNTSFIDTTPELFKPHIRDRGKTLYIGNVTVNGFGP 483
QY 505 VRPKDVAAPIDKLPNIKDLNLP----LPRGSRDLKQLGPAFAFARDLRODALAVTDTTFRD 560
DB 484 IKHRD--KPVYAEPLPKIPYGSQISGKTQILDAGKEGVVDWVKQOEVLTDITLRLD 541
QY 561 AHOSLLATVRSPFALKPAEAAVAKLTPELLSVEAWGATVDVAMRELFPDPWRDLRE 620
DB 542 AHOSLLATVRSKDI FOVADAMAHLLPNMFSFEMWGATFDVAIRFLEDPWRLETFLK 601
QY 621 AMPVNTOMLLRGNTVGYTPYDSCVRAVKEAASGVDFIRIFDALNDVSMRPAIDA 680
DB 602 QIPNVMQMLLRGANAVYKNYDNVIREFVKOSAGSDVDFVFDLSNLIKOMEVSIDA 661
QY 681 VLENTVAEVAAMAYSGLSDPNKLTLDYLYLKMAEIEVKSGAHLATKDMAGLLRPAA 740
DB 662 VREAG-KVSEAAICYTGIDIDDDTTRKYTIDYKDMAKELVAQTHILGDKMAGLLKPOA 720
QY 741 VTKLVTLRREFDLPHVHTHTDAGGOLATYFAAAQAGADAVDCASAPLSGTTTSPSLSA 800
DB 721 AYRLIGELKOTVDVPIHLHTDTSNGNIYTYAAAVSAGVDIVDVASSAMSGATSPSMTG 780
QY 801 IVAFAHTRRDTGLSLSEAVSDLEPYWEAVRGLYLPPESGTPGPTGRVYRHEIPGQLSNL 860
DB 781 LYGLVNGNRQTNLDAQNSQIINHVEDVRHYKDFDNALNSPQTEVYIHEMPGGQYTNL 840
QY 861 RAQATLGLDRFELIEDNTAAVNEMLGRPTKVTTPSSKVVGDIALHLVAGVDPADFAAD 920
DB 841 QOQAIAVGLGDRDEVKEMTYVNVQMGDIVKVTTPSSKVVGDIALHPVQNELSEEDVYEK 900
QY 921 POKYDIPDSVIAFLRGELNPPGMPPELPTRALEGRSEKAPLITEVPPEEQALHDADS 980
DB 901 GDTIDPDSVIEEFMGEIGQYGFPEKLOKVLKGT-----PLIDRPPALMEPVNFVDV 956
QY 981 K-----ERRNSNRLPLPKPTEPELEHRRRFGNTSALDREFFYGLVEGRETLI 1029
DB 957 KAELEKMGVEPTEKDVISILYKPVFLDYQDMINKYGDVTLVDLDTFTFKMRLGETIEV 1016
QY 1030 RLPDVRTPLVLRLDAISEPDDKGMNVNANVGQIRPMVRDSVESVTATAEKADSSNK 1089
DB 1017 ELEKGTI-LLIKLSNGEPTADTRIVYFELNGOPREINIQDMNVQSTVIARRKIDTTP 1075
QY 1090 GHVAAPFAG-VVTVVTAEGDEVKAGDAVATIEAMKMEATITASVDCKIERVVVPAATKVS 1148
DB 1076 EHVGMTGTSVIOVVVKGSDVKVKGDPPLITTEAMKMETTIQAPFDEGVSSIVVSGDGTIE 1135
QY 1149 GGLIVVVS 1157
```

```
DB 1136 SGDLLIEVN 1144
RESULT 13
Q92CW1 PRELIMINARY; PRT: 1146 AA.
AC Q92CW1;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE PYCA protein.
GN PYCA OR LIN1060.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BA Quiberon F., Berche P., Blocker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Denoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido F.,
GA Gautier L., Goebel W., Gomez-Lopez N., Hain F., Hauf J., Jackson D.,
RA Jones L.-M., Kaelst U., Klett J., Kuhn M., Kunst F., Nedjari H.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsted G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
EMBL; AL596167; CAC96291.1;
DR ListList; LIN01060;
DR InterPro; IPR000089; Biotin_lipoyl.
DR InterPro; IPR000901; CPSase.
DR InterPro; IPR000891; HMGL-like.
DR InterPro; IPR003379; PYC_OADA.
DR Pfam; PF02785; Biotin_carb_C; 1.
DR Pfam; PF00364; Biotin_lipoyl; 1.
DR Pfam; PF00289; CPSase_L_chain; 1.
DR Pfam; PF02786; CPSase_L_D2; 1.
DR Pfam; PF00682; HMGL-like; 1.
DR Pfam; PF02436; PYC_OADA; 1.
DR TIGRfams; TIGR01235; pyruv_carbox; 1.
DR PROSITE; PS00866; CPSASE_1; UNKNOWN_1.
DR PROSITE; PS00867; CPSASE_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1146 AA; 128035 MW; 3E23FFB4A289C60F CRC64;
Query Match 42.5%; Score 2491; DB 16; Length 1146;
Best Local Similarity 45.9%; Pred. No. 5.9e-125;
Matches 527; Conservative 183; Mismatches 409; Indels 30; Gaps 12;
QY 30 KILVANGETIAVAFRAALFETGAATVAIYPREDGRGSHRSFASFAVRIGTEGSPVKAYL 89
DB 5 KKLVANGETIAVYMRACTELKTKTVAIYSDPTGSHRYKSDYALVVGAKKPIDAYL 64
QY 90 DIDEIIGAARKVADAIYPGYGFLSENAQLARECAENGITFIGTPEVLDLTGDKSRVAT 149
DB 65 DIENIELAKESGADAIHPGYGFLSENFARFARCEQEGIIIFVGPCKSHLDMFGDKIKAKE 124
QY 150 AAKKAGLPVLAEST-PSKNIDDIIVKSGQOTPIFVAVAGGGRGMRFVSSPDELKLA 208
DB 125 QALLADIPVPGSNGPVAGIKEVEEFGKNGYPLMIKASLGSGGRGMRFVSKHEHVRESF 184
QY 209 TEASREAAAGDCSVYVERAVINPOHTEVOILGDRTEGVVHLVYERDCSLORRHOKVVEI 268
DB 185 ERASSEAKAAGNDEVYVEKCMVMPKHIEVOILGDTGNIVHLFERDCSLORRHOKVVEV 244
QY 269 APAQHLPELDRICADAVKFCRSIGYGAGTVEFLVDEKGNHVFIEEMNPRIQVEHTVTE 328
```

```

Db 245 APCNAITSLNRICDAAYKLMKNVDY INAGTVEFLV-EGDNFYFEVPRVQVEHTITE 303
QY 329 EYTEVDLVKAQRLAAGATLKLGLT--ODKIKTHGAALQCRITTEDPNNGFRPDGTGI 385
Db 304 MITGIDIVQSQFLIADGVALHDLVAIPKQEDIIHGSIAIQSRITTEDPLNNFMPDGTGRV 363
QY 386 TAYRPGGAGVRLD-GAALGGEITAHFDSMLVKMTCRGSDFTAVARAQALAEFTVSG 444
Db 364 DTYRSTGGFVRLDAGNGFGIVVTFYDLSLVKLTGWTTEQATRKRRNLIERIRG 423
QY 445 VATNTGFLRALLREEDFTSKRIATGIGDHPHLLQAPPADDEGRILDYADVYVKNKPH 504
Db 424 VRTNIPFLNVRHPDFAFSGNTSPIDTTPLEFKPHIRDRGKTLRIGNVTVNGFPG 483
QY 505 VRPKDVAAPIDKLPNIKDL-P---LPRGSRDLKQLGPAARFDRLEQDALAVTDTTFRD 560
Db 484 IKHRD--KPYAEPRLPKIPYGSQIAPGTQKILDAKPGCVVDWKKQBEVLDTFTLKD 541
QY 561 AHQSLLATRVRSFALKPAEAVALKTPPELLSVSEAWGATYDVAMRFLFEDPMDRLDELRE 620
Db 542 AHQSLLATRVRSKDIQIADAMAHLLPNMFSEFMWGATFDVAYRFLNEDPWVRLTLRK 601
QY 621 AMPNVNIQMLRGNVTGTPYDPSVCRAPVKEAASSGVVDIERIFDALNDVSOMRPAIDA 680
Db 602 QIPNVMFQMLRGANAAGYKNTPDNVIREFVKQAQSGVDVFRVFDLSLNKIKGMEVSIDA 661
QY 681 VLENTAVAEVAMAYSGDLSDNKELYLDYLYKMAEEIVKSCAHLIAKDMAGLLRPAA 740
Db 662 VREAG-KVYEATICYGDDIDDTRTKTYIDYKDMAKELVAOGTHILGDKMAGLLKPOA 720
QY 741 VTKLVTALRREFDLPVHVHTHTAGQLATYFAAAQAGADAVDAGSAPLSGTTSQPSLSA 800
Db 721 AYRLIGELKDTVDVPIHLHTHTDSNGIYVYAAAVSAGVDIVDVASSAMSGATSQPSMTG 780
QY 801 IVAAFATRTDGLSLEAVSDLEPTWEAVRGLYLPFESGTGPTGRVYRHEITPGGOLSNL 860
Db 781 LYYGLVNGNRQTNLDQNSQIINHMYEDVRHYKDFDNALNSPQTEVYIHEMPGGQYTNL 840
QY 861 RAQATALGLADRFELLEDNYAAVNMELGRPTKVTSPSKVVGDALHLVAGVDPADFAAD 920
Db 841 QQAATAVGLGDRDEKEMVTYVQNGFDIVKVTSPSKVVGDLALFWQNELTEEDVYEK 900
QY 921 POKYDIPDSVIAFLRGELOPNPGWPEPLRTRALEGRSECKAPLTEVP-----EEE 971
Db 901 GDTIDPDSVIEFFMGEIGOPYGGPEKLOKLKGR-----PLADRPCALMEPVNFAEV 956
QY 972 OAHLDADDSKE--RRNSLNRLFPKPTFEFLHRRRFGNTSALDDREFFYGLVEGRETLI 1029
Db 957 KAELEKMGYEPSEKDVISYILYKVFLDYQEMISKYGDVTVLDTPTFYKMRGLGETIEV 1016
QY 1030 RLDPDVRTPLVRLDAISEPDDKGRNVVANVNGOIRPMVRDRSVESVTATAEKADSSNK 1089
Db 1017 ELEKGI-LIKLNSICEPIADGTRVYFELNGQPREINIQDMNVOSTVIARRKIDTTNP 1075
QY 1090 GHVAAPFAG-VYTVTVAEDEVKAGDAVAIIIEAMKMEATITASVDGKIERVVVPAATKVE 1148
Db 1076 EHVGMTMTSGVIVVVKKGDSVKKGPPLLITEAMKMETTIQAPFDGEVSSIVVSDGDTIE 1135
QY 1149 GGDLIWVS 1157
Db 1136 SGLLIEVN 1144

```

RESULT 14

Q97FR7 ID Q97FR7 PRELIMINARY; PRT: 1144 AA.  
AC Q97FR7  
DT 01-OCT-2001 (TReMBLrel. 18, Created)  
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Pyruvate carboxylase, PYKA.  
GN CAC2660.

OS Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;  
OC Clostridiales; Clostridiaceae; Clostridium.  
OX NCBI\_TaxID=1488;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=21359325; PubMed=11466286;  
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I., Daly M.J.,  
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Tatusov R.L.,  
RA Bennett G.N., Koonin E.V., Smith D.R.;  
RT "Genome sequence and comparative analysis of the solvent-producing  
RT bacterium Clostridium acetobutylicum.";  
RL J. Bacteriol. 183:4823-4838(2001).  
DR EMBL: AE007763; RAK80607.1; -;  
DR InterPro: IPR001882; Biotin\_attach.  
DR InterPro: IPR000089; Biotin\_lipoyl.  
DR InterPro: IPR000901; CPSase.  
DR InterPro: IPR000891; HMGL-like.  
DR InterPro: IPR003141; PHP\_N.  
DR InterPro: IPR003379; PYC\_OADA.  
DR Pfam: PF02785; Biotin\_carb\_C; 1.  
DR Pfam: PF00364; biotin\_lipoyl; 1.  
DR Pfam: PF00289; CPSase\_L\_chain; 1.  
DR Pfam: PF02786; CPSase\_L\_D2; 1.  
DR Pfam: PF00682; HMGL-like; 1.  
DR Pfam: PF02436; PYC\_OADA; 1.  
DR PRINTS: PR00098; CPSASE.  
DR SMART: SM00481; POLIHIAC; 1.  
DR TIGRFAMS: TIGR01235; pyruv\_carbox; 1.  
DR PROSITE: PS00188; BIOTIN; UNKNOWN\_1.  
DR PROSITE: PS00866; CPSASE\_1; 1.  
DR PROSITE: PS00867; CPSASE\_2; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 1144 AA; 127709 MW; 519FA29A8008F326 CRC64;

Query Match 42.5%; Score 2490.5; DB 16; Length 1144;  
Best Local Similarity 45.2%; Pred. No. 6.2e-125;  
Matches 519; Conservative 204; Mismatches 392; Indels 33; Gaps 14;

QY 29 FKKILVANGETAVRAFALETCATVAIYPREDGRSHRSFASAVRITGEGSPVKAY 88  
Db 5 FKKVLVANGETAIRIFRACHELGITVAIYSEEDLALFRKADESYLIGQNRKGPVDAY 64  
QY 89 LDDEITIGAAKKVKADAIYPGYGFLSENAOLARECAENGITFTGPTPEVLDTLGDKSRAV 148  
Db 65 LNDEIINIALKGVDAIHPGYGFLSENSEFSRRCTEAGIEFTGPTGDMMDKLGDKINSK 124  
QY 149 TAAKAGLPVL-AESTPSKNIDDIVKSAEQTYPIFVAVAGGGGRGMRVSSPDRLKL 207  
Db 125 LAAGAAGVKTIPGVEKPIETEQAIEFARTCGYPVMVKAAGGGGRMRIVEREDLIAA 184  
QY 208 ATEASREAAAFGDSGVYVERAVINPOHIEVLGORTGEVHLYERDCSLQRHOKVVE 267  
Db 185 CRSASEAKKAFGIEDIEFKYLEGPKHIEVQVLGKYGNIVHLYERDCSVQRHOKVIE 244  
QY 268 IAPQAHLDPDLRICADAVKFCRSIGYOGAGTVEFLVDEKGNHVRITEMPRIOVEHTVT 327  
Db 245 LTPAVSMSEKRLCEIDALKIARSIGYSAGTLEFLDKHGNHYFTEMPRVOVEHTIT 304  
QY 328 EEVTEVDLVKAQMLAAGATLK--ELGL-TQDKIKTHGAALQCRITTEDPNNGFRPDGTG 384  
Db 305 EMTVDIGIVSQILIAEGYKLSNPEVINSQEDIHVNGYAIQCRITTEDPSNSFAPDTGK 364  
QY 385 ITAYRSPGGAGVRLDGA-AQLGGEITAHFDSMLVKMTCRGSDFTAVARAQALAEFTVS 443  
Db 365 IDVYRTGSGFIRLDGGNGFTGAVISPYDLSLVKSTWSRSTFEDAIRKAIRAIKETYIS 424  
QY 444 GVATNIGFLRALLREEDFTSKRIATGFIGDHPHLLQAPPADDEGRILDYADVYVKNKPH 503  
Db 425 GVKTNIDFLNVLNHNHETFRKGLCDTNFIANNPELFEITPRIDELRVLFKIGERKVVNETH 484



Db 647 FDSLNVDMQKVANEAVQEAG-KISEGTCYTGDIINPERSNIYTTLEYVVKLAKELEREG 705  
QY 724 AHILAIKDMAGLLRPAAVTKLVTALRREFDLPVHVHTDHTAGGOLATYFAAAQAGADAVD 783  
Db 706 FHILAIKDMAGLLKPAANAELIGELKASANDLPJHLHTDHTSGNGLTTYKQADAGVDIID 765  
QY 784 GASAPLSGTTSQPSLSAISAAFAHTRDRTGLSLEAVSDLEPYWEAVRGYLPFESGTPGP 843  
Db 766 TAVASMSGLTSQPSANSLYALNGFPHLRTDIEGMESLSHYWSTVRYTYSDFESDIKSP 825  
QY 844 TGRVYRHEIPGGOLSNLRAQATAGLADREFELIEDNYAAVNEMLGRPTKVTSSKVVGD 903  
Db 826 NTEIYQHMPGGQYSLNLSQAQSLGLGERFDEVKDMYRRVNFGLGDIVKVTSSKVVGM 885  
QY 904 ALHLVGAGVDPADFAADPKYDIPDSVIAFLRGELGNPPGWPPELRTALERSEKAP 963  
Db 886 ALYMWQNDLDEQSVITDGYKLDFEPESVSVFFKEIGEQPVNGFNKDLQAVILKQE----A 941  
QY 964 LTEVPEEQAHLDADDSKE-----RRNSLNRLLPKPTPEFLEHRRRFGNTSAL 1012  
Db 942 LTARPGYLEPVDPEKVRLEEEEOQGPVTEQDIISVLYPKVVEQYIQTRNQYGNLSLL 1001  
QY 1013 DDEFFYGLVEGRETLIRLPDVRTPLLVRLDAISEPDDKGMNVVANVNGQIRPMRVDR 1072  
Db 1002 DTPTFFFGMRNGETVEIEI-DKGKRLIIKLETISEPDENGNTIYYAMNGQARRIYKDE 1060  
QY 1073 SVESVTATAEKADSSNKGHVAAPFAGVVT-VTVAEGDEVKAGDAVAIIEMKMEATITAS 1131  
Db 1061 NVHTNANVPKADKSNPSHIGAQMPGSVTEVKVSGETVKANOPLLITEAMKMETTIQAP 1120  
QY 1132 VDGKIERVVVPAATKYEGGDLIVV 1156  
Db 1121 FDGVIKQVTNVNGDTIATGDLLEI 1145

Search completed: March 26, 2003, 19:02:39  
Job time : 117 secs

